

05-24-00

box Seq.

A

Please type a plus sign (+) inside this box → ☐

Approved for use through 09/30/2000. OMB 0651-0032
 Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 C.F.R. § 1.53(b))

Attorney Docket No. 601-1-057N
 First Inventor or Application Identifier William G. Johnson
 Title METHODS FOR DIAGNOSING, ...
 Express Mail Label No. EL629423850US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. ☐ Fee Transmittal Form (e.g., PTO/SB/17)
 (Submit an original and a duplicate for fee processing)
2. ☒ Specification [Total Pages 109]
 (preferred arrangement set forth below)
- Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 5]
4. Oath or Declaration unexecuted [Total Pages 3]
- a. ☐ Newly executed (original or copy)
- b. ☐ Copy from a prior application (37 C.F.R. § 1.63(d))
 (for continuation/divisional with Box 16 completed)
1. ☐ DELETION OF INVENTOR(S)
 Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).

NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).

16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:
☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: _____

Prior application information: Examiner _____

Group / Art Unit: _____

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or ☐ Correspondence address below

Name David A. Jackson
 Klauber & Jackson
 Address Continental Plaza
 411 Hackensack Avenue
 City Hackensack State New Jersey Zip Code 07601
 Country U.S.A. Telephone 1(201)487-5800 Fax 1(201)343-1684

Name (Print/Type) Michael D. Davis Registration No. (Attorney/Agent) 39,161
 Signature Michael D. Davis Date May 23, 2000

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

05/23/00

J-6813 U.S. PTO

00577266-652300

05/23/00

09/57/246

METHODS FOR DIAGNOSING, PREVENTING, AND TREATING DEVELOPMENTAL DISORDERS DUE TO A COMBINATION OF GENETIC AND ENVIRONMENTAL FACTORS

CROSS-REFERENCE TO RELATED APPLICATIONS

- 5 The present application is a non-provisional application claiming the priority of copending provisional U.S. Serial No. 60/136,198 filed May 25, 1999, the disclosure of which is hereby incorporated by reference in its entirety. Applicants claim the benefits of this application under 35 U.S.C. §119(e).

FIELD OF THE INVENTION

- 10 The invention relates generally to novel methods of diagnosing, preventing, and treating specific diseases which are caused by a combination of genetic and environmental factors. One such disease exemplified is schizophrenia.

BACKGROUND OF THE INVENTION

- The term "schizophrenia" was introduced by Bleuler in the beginning of this century
15 to encompass a dissociation or disruption of thought processes, along with a dichotomy among thought, emotion, and behavior [Bleuler, *Translation J. Zinkin*, New York: International University Press (1950)]. The current definition of schizophrenia includes a break with reality that is usually manifested as hallucinations, delusions, or disruption in thought processes [Carpenter *et al.*, *Medical*
20 *Progress*, **330**:681-690 (1994)]. At present the nationally accepted definition for the diagnosis of schizophrenia is contained in Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition, Washington, D.C (1994): American Psychiatric Association, hereby incorporated by reference in its entirety.

- Schizophrenia is a clinical syndrome that has a profound influence on public health.
25 The symptoms for schizophrenia begin early in life, and continues for most patients throughout their lives. An estimate of the direct and indirect costs of schizophrenia was thirty-three billion dollars for 1990 in the United States alone [Carpenter *et al.*,

1994, *supra*]. Indeed, one of every forty dollars spent for total health care expenditures in the United States is spent on treating schizophrenia [Rupp *et al.*, *Psychiatric Clin. North Am.*, **16**:413-423 (1993)]. Furthermore, estimates have been made suggesting that up to 50% of the homeless American population is schizophrenic [Bachrach, In: *Treating the Homeless Mentally Ill*, Washington, D.C., American Psychiatric Press, 13-40, Lamb *et al.* ed. (1992)].

- The genetic factors in schizophrenia, though clearly documented to be present, are not simple [Carpenter and Buchanan, *N. Engl. J. Med.*, **330**:681-689 (1994); Gottesman, *Clin. Genet.*, **46**:116-123 (1994)]. Schizophrenia is, at least in part, a neurodevelopmental disorder, a birth defect in which the brain has been subtly damaged during development [Carpenter and Buchanan, *N. Engl. J. Med.*, **330**:681-689 (1994); Weinberger, *Arch. Gen. Psychiatry*, **44**:660-669 (1987); Brixey *et al.*, *J. Clin. Psychol.*, **49**:447-456 (1993)]. Evidence of this damage is seen both at autopsy [Kovelman and Scheibel, *Biol. Psychiatry*, **19**:1601-1621 (1984); Bogerts *et al.*, *Arch. Gen. Psychiatry*, **42**:784-791 (1985); Jakob and Beckman, *J. Neural Transm.*, **65**:303-326 (1986); Brown *et al.*, *Arch. Gen. Psychiatry*, **43**:36-42 (1986); Benes and Bird, *Arch. Gen. Psychiatry*, **44**:608-616 (1987); Colter *et al.*, *Arch. Gen. Psychiatry*, **44**:1023 (1987); Altshuler *et al.*, *Arch. Gen. Psychiatry*, **47**:1029-1034 (1990); Pakkenberg, *Schizophr. Res.*, **7**:95-100 (1992); Bogerts, *Schizophr. Bull.*, **19**:431-445 (1993); Shapiro, *Schizophr. Res.*, **10**:187-239 (1993)] and by neuroimaging [Jeste *et al.*, *Br. J. Psychiatry*, **153**:444-459 (1988); Suddath *et al.*, *Am. J. Psychiatry*, **146**:464-472 (1989); Suddath *et al.*, *N. Engl. J. Med.*, **322**:789-794 (1990); DeLisi *et al.*, *Biol. Psychiatry*, **29**:159-175 (1991); Breier *et al.*, *Arch. Gen. Psychiatry*, **49**:921-926 (1992); O'Callaghan *et al.*, *J. R. Soc. Med.*, **85**:227-231 (1992); Bogerts *et al.*, *Biol. Psychiatry*, **33**:236-246 (1993); Andreasen *et al.*, *Science*, **266**:294-298 (1994)]. The pattern of this brain damage and the presence of minor congenital abnormalities point to an insult occurring during the second trimester of fetal development [Bracha *et al.*, *Biol. Psychiatry*, **30**:719-725 (1991); Bracha *et al.*, *Am. J. Psychiatry*, **149**:1355-1361 (1992); Green *et al.*, *Psychiatry Res.*, **53**:119-127 (1994)].
- Epidemiological studies have documented a season-of-birth effect by which schizophrenics are more frequently born during winter and early spring than during other seasons [Boyd *et al.*, *Schizophr. Bull.*, **12**:173-186 (1986); Kendell and Adams,

Br. J. Psychiatry, **158**:758-763 (1991); O'Callaghan *et al.*, Br. J. Psychiatry, **158**:764-769 (1991)]. Also, individuals exposed to an influenza epidemic [Mednick *et al.*, Arch. Gen. Psychiatry, **45**:189-192 (1988); Barr *et al.*, Arch. Gen. Psychiatry, **47**:869-874 (1990); O'Callaghan *et al.*, Lancet, **337**:1248-1250 (1991); Murray *et al.*, J. Psychiatr. Res., **26**:225-235 (1992); Adams *et al.*, Br. J. Psychiatry, **163**:522-534 (1993)] or famine [Susser and Lin, Arch. Gen. Psychiatry, **49**:983-988 (1992)] during their second trimester of fetal development have increased risk of later developing schizophrenia, according to some studies but not others [Kendell, Arch. Gen. Psychiatry, **46**:878-882 (1989); Crow and Done, Br. J. Psychiatry, **161**:390-393 (1992)]. This has suggested that an environmental effect such as dietary deficiency, virus infection [Kirch, Schizophr. Bull., **19**:355-370 (1993)], vitamin deficiency, or effect of cold weather may be acting during fetal development.

Linkage mapping studies in schizophrenia have been difficult. Recently, some studies [Straub *et al.*, Nature Genet., **11**:287-293 (1995); Schwab *et al.*, Nature Genet., **11**:325-327 (1995); Moises *et al.*, Nature Genet., **11**:321-324 (1995)] have supported a gene locus on chromosome 6 (6p24-22, near the HLA region) as having an effect in schizophrenia; other studies gave little or no support to a marker in this region [Wang *et al.*, Nature Genet., **10**:41-46 (1995); Mowry *et al.*, Nature Genet., **11**:233-234 (1995); Gurling *et al.*, Nature Genet., **11**:234-235 (1995); Antonarakis *et al.*, Nature Genet., **11**:235-236 (1995)]. At best this locus appeared to be involved in only about 15-30% of families [Straub *et al.*, 1995, *supra*]. Also, some evidence for loci on chromosomes 3 [Pulver *et al.*, Am. J. Med. Genet., **60**:252-260 (1995), 8 [Pulver *et al.*, Am. J. Med. Genet., **60**:252-260 (1995); Kendler *et al.*, Am. J. Psych. **153**:1534-1540 (1996), 9 [Coon *et al.*, Biol. Psychiatry, **34**:277-289 (1993); Moises *et al.*, Nature Genet., **11**:321-324 (1995)] and 22 [Coon *et al.*, Am. J. Med. Genet., **54**:72-79 (1994); Pulver *et al.*, Am. J. Med. Genet., **54**:3-43 (1994)] have been reported. In addition, two polymorphic markers very close to the gene encoding dihydrofolate reductase (DHFR) on chromosome 5q, D5S76 and D5S39, gave very high lod scores (as high as 6.49, *i.e.* odds of about 3 million to one in favor of genetic linkage versus chance occurrence) in 7 British and Icelandic schizophrenia families studied [Schwab *et al.*, Nat. Genet. **11**:325-327 (1997); Straub *et al.*, Molec

Psychiatr. 2:148-155 (1997)]. However, this result could not be confirmed in studies of numerous other families.

There could be several reasons for this difficulty. First, there may be more than one gene involved, (locus heterogeneity). Second, the genetic factor(s) may be common in the population (high disease allele frequency), thus diminishing the power of linkage studies [Terwilliger and Ott, *Handbook of Human Genetic Linkage*, Baltimore: Johns Hopkins Univ. Pr., 181 (1994)]. Third, the correct genetic model may be unknown [Owen, *Psychol. Med.*, **22**:289-293 (1992)]. Any or all of these factors could diminish the power of a linkage study sufficiently to make success very difficult [Terwilliger and Ott, 1994, *supra*].

Thus the current (developmental) model for schizophrenia is that genetic and environmental factors cause brain damage in a fetus that later develops schizophrenia. However, the genetic and environmental factors have not been identified. Also, extensive linkage and association studies have failed to identify genes determining schizophrenia.

Indeed, schizophrenia appears to be just one of a family of developmental disorders whose cause has not been identified. Other such developmental disorders are defined by the Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition, Washington, D.C (1994) and include: Tourette Syndrome which is identical to Tourette's Disorder and is a subcategory of Tic Disorders; Bipolar Disorder which is identical with Bipolar I Disorder or Bipolar II disorder; Autism which is identical with Autistic Disorder which is a subcategory of Pervasive Developmental Disorders; Conduct disorder which is a subcategory of Attention-Deficit and Disruptive Behavioral Disorders; Attention-Deficit Hyperactivity Disorder which is identical to Attention-Deficit/Hyperactivity Disorder and to Attention-Deficit/Hyperactivity Disorder NOS (not otherwise specified) which is also a subcategory of Attention-Deficit and Disruptive Behavioral Disorders; Obsessive-Compulsive Disorder which is a subtype of Anxiety Disorders; Chronic Multiple Tics Syndrome which is identical to Chronic Motor or Vocal Tic Disorder which is a subtype of Tic Disorders; and Learning Disorders.

In addition Spina bifida is a developmental disorder. Spina bifida is a form of neural tube defect in which neural elements (spinal nerves or spinal chord) or coverings of the brain and spinal chord (dura mater, arachnoid mater) herniate through a midline defect into a cystic cavity covered completely or partially by skin.

- 5 Therefore, there is a need for new methods of diagnosing individuals susceptible to developing a developmental disorder. In addition, there is a need for methods of identifying individuals susceptible to having offspring that develop a developmental disorder. Finally, there is a need for a method of treating such susceptible individuals in order to prevent and/or ameliorate the symptoms due to and/or associated with the
- 10 developmental disorder.

The citations of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application.

SUMMARY OF THE INVENTION

- The present invention provides methods of diagnosing, preventing and/or treating
- 15 specific developmental disorders. Towards this end the present invention provides methods of identifying an individual as being genetically or environmentally susceptible for developing or having a developmental disorder or for having offspring that develop the developmental disorder. Such a developmental disorder can be schizophrenia, spina bifida cystica, Tourette's syndrome, bipolar illness, autism,
- 20 conduct disorders, attention deficit disorder, obsessive compulsive disorder, chronic multiple tic syndrome and learning disorders such as dyslexia. In addition, any of the methods provided herein for identifying an individual as being genetically and/or environmentally susceptible for having or developing a developmental disorder or for having offspring that develop the developmental disorder can also be used in
- 25 diagnosing the individual, preferably in conjunction with a clinical diagnosis.

Therefore, the present invention provides methods of identifying an individual as being genetically susceptible for having or developing a developmental disorder.

The present invention further provides methods of identifying an individual as being genetically susceptible for having offspring that are susceptible for developing a developmental disorder. Methods of identifying an individual as being susceptible due to environmental factors for having or developing a developmental disorder are also provided. In addition, the present invention provides methods of identifying an individual as being susceptible of having offspring that are susceptible for developing a developmental disorder. The present invention also provides methods of identifying an individual as being susceptible for having or developing a developmental disorder due to both environmental and genetic factors. The present invention further provides methods of identifying an individual as being susceptible for having offspring that are susceptible for developing a developmental disorder

The present invention therefore provides methods for compiling genetic reference datasets, environmental reference datasets and/or genetic and environmental reference datasets for use in determining a predicted probability for an individual of having a susceptibility for having or developing a developmental disorder, or for having offspring that develop a developmental disorder.

In one aspect of the invention, the present invention provides methods that comprise generating a genetic reference dataset for use in determining the predicted probability of an individual for having a susceptibility for having or developing a developmental disorder due to genetic factors, or for having offspring that develop a developmental disorder due to genetic factors.

One such embodiment comprises collecting a biological sample from a human subject. The human subject can be a diagnostic proband, a blood relative of the diagnostic proband, an affected proband, a blood relative of the affected proband, a control proband, and/or a blood relative of the control proband. The biological sample contains nucleic acids and/or proteins from the human subject. The nucleic acids and/or proteins from the biological sample are then analyzed resulting in a partial or full genotype for the alleles of the genes involved in folate, pyridoxine, and/or cobalamin metabolism. The partial or full genotype then forms a dataset of genetic explanatory variables for the human subject. The dataset of genetic

explanatory variables is then compiled from multiple human subjects into a genetic reference dataset. Such compilations are exemplified in the Detailed Description and Examples below.

In another aspect, the present invention provides a method that comprises generating a genetic and environmental reference dataset for use in determining the predicted probability of an individual for having a susceptibility for having or developing a developmental disorder due to genetic factors and environmental factors, or for having offspring that develop a developmental disorder due to genetic factors and environmental factors. One such embodiment comprises obtaining dietary and epidemiological information for environmental explanatory variables for the human subjects and combining the environmental explanatory variables with a genetic reference dataset for the human subjects as described above.

In another aspect, the present invention provides an environmental reference dataset for use in the determination of the predicted probability for an individual for having a susceptibility for having or developing a developmental disorder due to environmental factors, or for having offspring that develop a developmental disorder due to environmental factors. One such embodiment comprises obtaining dietary and epidemiological information for environmental explanatory variables for a human subject. The human subject can be a diagnostic proband, a blood relative of the diagnostic proband, an affected proband, a blood relative of the affected proband, a control proband, or a blood relative of the control proband. The dataset of environmental explanatory variables is then compiled from multiple human subjects into an environmental reference dataset for the human subjects.

The developmental disorder forming the basis of the reference datasets of the present invention can be schizophrenia, or spina bifida cystica, or Tourette's syndrome, or dyslexia, or conduct disorder, or attention-deficit hyperactivity disorder, or bipolar illness, or autism, or chronic multiple tic syndrome or obsessive-compulsive disorder, or like disorders. A blood relative is preferably the mother of the individual, a sibling, the father or a grandparent of the individual. When the reference dataset is for use in the determination of the predicted probability for an individual of having a

susceptibility for having offspring that develop a developmental disorder, the individual is preferably a pregnant woman. The reference datasets of the present invention are themselves part of the present invention.

The present invention further provides methods of estimating the genetic
5 susceptibility of an individual to have or to develop a developmental disorder, or to have offspring that develop a developmental disorder. In one such embodiment the method comprises collecting a biological sample from a participant (or participants) who is either the individual or a blood relative of the individual. The biological
10 sample contains nucleic acids and/or proteins of the participant. The analysis of the nucleic acids and/or proteins from the biological sample yield a partial or full genotype for the alleles of the genes involved in folate, pyridoxine, and/or cobalamin metabolism. The partial or full genotype forms a dataset of genetic explanatory variables for the participants. The dataset of genetic explanatory variables obtained
15 are added to a genetic reference dataset forming a combined genetic dataset. A model is then formulated comprising the genetic explanatory variables obtained from the participants and the combined genetic dataset is analyzed. A predicted probability for the individual for having and/or developing a developmental disorder and/or having
20 offspring that develop a developmental disorder is then determined. The genetic susceptibility of an individual to have or to develop a developmental disorder and/or have offspring that develop a developmental disorder is estimated. In a preferred embodiment, analyzing the combined genetic dataset is performed by binary linear regression. In a more preferred embodiment, the binary linear regression is performed with the SAS system. In another preferred embodiment, the model is modified by adding or subtracting one or more genetic explanatory variables and the
25 combined genetic dataset is re-analyzed, preferably by binary logistic regression. In this case a model is chosen that best fits the data. This can be accomplished by testing the model for goodness of fit.

The present invention also provides methods of estimating the genetic and
environmental susceptibility of an individual to have or to develop a developmental
30 disorder and/or for having offspring that develop a developmental disorder. One such embodiment comprises collecting a biological sample from one or more participants.

- Again, the participant is either the individual or a blood relative of the individual. The biological sample contains nucleic acids and/or proteins of the participant. The nucleic acids and/or proteins from the biological sample are analyzed resulting in a partial or full genotype for the alleles of the genes involved in folate, pyridoxine, and/or cobalamin metabolism. The partial or full genotype forms a dataset of genetic explanatory variables for the participant. Dietary and epidemiological information for environmental explanatory variables for the participant(s) are also obtained which are used to form a dataset of environmental explanatory variables for the participant(s). The datasets of genetic explanatory variables and the dataset of environmental explanatory variables are added to a genetic and environmental reference dataset forming a combined genetic and environmental dataset. A model is formulated comprising the genetic and environmental explanatory variables obtained from the participant(s). The combined genetic and environmental dataset is then analyzed and a predicted probability for the individual for having and/or developing a developmental disorder and/or for having offspring that develop a developmental disorder is determined. The genetic and environmental susceptibility of an individual to have or to develop a developmental disorder and/or have offspring that develop a developmental disorder is estimated. In a preferred embodiment, analyzing the combined genetic and environmental dataset is performed by binary linear regression.
- In a more preferred embodiment the binary linear regression is performed with the SAS system. In another preferred embodiment the model is modified by adding or subtracting one or more genetic and/or environmental explanatory variables and the combined genetic and environmental dataset is re-analyzed preferably, by binary logistic regression. In this case a model is chosen that best fits the data. This can be accomplished by testing the model for goodness of fit.

For any of these methods, the developmental disorder can be schizophrenia, spina bifida cystica, Tourette's syndrome, bipolar illness, autism, conduct disorder, attention deficit hyperactivity disorder, obsessive compulsive disorder, chronic multiple tic syndrome and learning disorders such as dyslexia.

- In a particular embodiment, the individual is suspected of being genetically susceptible of having or for developing the developmental disorder and/or of being

genetically susceptible of having offspring that develop the developmental disorder. In a preferred embodiment of this type, the individual is suspected of being genetically susceptible for having or for developing the developmental disorder and/or of being genetically susceptible of having offspring that develop the

5 developmental disorder because a blood relative has the developmental disorder. In one such embodiment the blood relative is a parent, a sibling, or a grandparent. In a preferred embodiment the blood relative is the mother of the individual. In a particular embodiment in which the individual is suspected of being genetically susceptible of having offspring that develop the developmental disorder, the

10 individual is a pregnant woman. In another such embodiment the individual is the mate of the pregnant woman. In a particular embodiment exemplified below, the developmental disorder is schizophrenia.

Since the availability of the data regarding the genetic and environmental explanatory factors can vary in separate determinations, variations in the explanatory factors used

15 is clearly envisioned by the present invention.

The present invention further provides methods of lowering the risk of a pregnant woman to have a child that will develop a developmental disorder. One such embodiment comprises administering methylfolate, cobalamin or pyridoxine to the pregnant woman and/or fetus, which lowers the risk of the pregnant woman to give

20 birth to a child with a developmental disorder. In a particular embodiment of this type, the pregnant woman had been previously determined to be susceptible of having offspring that develop a developmental disorder by a method disclosed herein. The present invention further provides a method of determining if any treatment is advisable for a pregnant woman that is genetically susceptible to having offspring that

25 develop a developmental disorder which comprises determining the concentration of a risk factor from a tissue sample or body fluid from the pregnant woman. When the concentration of the risk factor is statistically above or below an accepted normal range, treatment is advisable.

The present invention further provides methods of determining if any treatment is

30 advisable for a pregnant woman who has been determined to be susceptible to having

offspring that develop a developmental disorder. One such embodiment comprises determining the concentration of a risk factor from a tissue sample or body fluid from the pregnant woman. When the concentration of the risk factor is statistically above or below an accepted normal range, treatment is advisable. In a particular
5 embodiment of this type, the pregnant woman had been previously determined to be susceptible of having offspring that develop a developmental disorder by a method disclosed herein.

Methods of monitoring the effect of the administration of methylfolate, cobalamin or pyridoxine to the pregnant woman who has been determined to be susceptible to
10 having offspring that develop a developmental disorder are also included in the present invention. One such embodiment comprises determining the concentration of a risk factor from a tissue sample or body fluid from the pregnant woman. When the concentration of the risk factor is statistically within an accepted normal range, the treatment is deemed effective. In a particular embodiment of this type, the pregnant
15 woman had been previously determined to be susceptible of having offspring that develop a developmental disorder by a method disclosed herein. The risk factor can be any substance and/or metabolite linked to folate and/or cobalamin and/or pyridoxine metabolism. In one embodiment, the risk factor is homocysteine. In yet another embodiment, the risk factor is folate. In still another embodiment, the risk
20 factor is cobalamin.

The present invention also provides a method of treating an asymptomatic individual determined to be susceptible for developing a developmental disorder comprising administering methylfolate, cobalamin and/or pyridoxine. In a particular embodiment of this type, the asymptomatic individual had been previously determined to be
25 susceptible of developing a developmental disorder by a method disclosed herein.

The DNA samples from the persons tested may be obtained from any source including blood, a tissue sample, amniotic fluid, a chorionic *villus* sampling, cerebrospinal fluid, and urine.

- The present invention includes but is not limited to the examples of proteins encoded by genes involved in folate, cobalamin and pyridoxine metabolism compiled in Tables 2-7 in the Detailed Description of the Invention, below. For certain genes nucleic acid and/or amino acid sequence data is also provided. These genes and
- 5 related sequence data are solely intended as examples of genes that are suitable to be used in the methods described herein. Such sequence data can be used for carrying out the genetic analysis of the present invention. However, the present invention is not intended to be limited in any way to such lists of proteins or the related sequence data.
- 10 It is further contemplated by the present invention to provide methods that include the testing for a genetic mutations in individual genes involved in folate and cobalamin metabolism and/or in individual combinations of such genes (*e.g.*, methylenetetrahydrofolate reductase gene and methionine synthase). In addition, all possible combinatorials, and permutations of such genes including a constellation
- 15 comprising all of the genes involved in folate, pyridoxine, and cobalamin metabolism is envisioned by the present invention. Alternatively, a constellation of genes in which any one or more genes can be excluded from those tested is also contemplated by the present invention (for example, a given constellation of genes can include genes encoding all of the proteins in Table 2 and 4 except the folate receptor 2-like
- 20 protein). Thus all of such possible constellations are envisioned by, and are therefore part of the present invention.

- The present invention also provides DNA polymorphisms that can be used as genetic explanatory factors in the present invention. One such embodiment is a nucleic acid encoding a genetic variant of human dihydrofolate reductase comprising a nucleotide
- 25 sequence having a 19 base-pair deletion spanning nucleotides 540 to 558 of the nucleotide sequence of SEQ ID NO:41. In a preferred embodiment the nucleic acid has the nucleotide sequence of SEQ ID NO:42.

- The present invention also includes primers. One such embodiment is a PCR primer that can be used to distinguish SEQ ID NO:42 from SEQ ID NO:41. Another
- 30 embodiment is a PCR primer that can be used to distinguish SEQ ID NO:42 from

SEQ ID NO:45. These primers are useful for identifying the 19 base-pair deletion spanning nucleotides 540 to 558 of the nucleotide sequence of SEQ ID NO:41 (see Example 2). In a particular embodiment, the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of SEQ ID NO:41. In another embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of the complementary strand of SEQ ID NO:41. In still another embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of SEQ ID NO:42. In yet another embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of the complementary strand of SEQ ID NO:42. In still another embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of SEQ ID NO:45. In yet another embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the nucleotide sequence of the complementary strand of SEQ ID NO:45.

In a particular embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from nucleotides 350 to 530 of SEQ ID NO:41. In a preferred embodiment of this type, the PCR primer has the nucleotide sequence of CTAAACTGCATCGTCGCTGTG (SEQ ID NO:38). In another particular embodiment the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the complementary strand of nucleotides 550 to 850 of SEQ ID NO:41. In preferred embodiment of this type, the PCR primer comprises 8 to 100 and preferably 10 to 50 consecutive nucleotides from the complementary strand of nucleotides 570 to 690 of SEQ ID NO:41. In a particular embodiment, the PCR primer has the nucleotide sequence of AAAAGGGGAATCCAGTCGG (SEQ ID NO:39).

The present invention also provides a nucleic acid that hybridizes under standard hybridization conditions to the nucleotide sequence ACCTGGGCGGGACGCGCCA (SEQ ID NO:40). In another embodiment the nucleic acid hybridizes under standard hybridization conditions to the nucleotide sequence complementary to SEQ ID NO:40. In yet another embodiment the nucleic acid hybridizes under standard

hybridization conditions to the nucleotide sequence ACCTGGGCGGGACGCGCC (SEQ ID NO:46). In yet another embodiment the nucleic acid hybridizes under standard hybridization conditions to the nucleotide sequence complementary to SEQ ID NO:46. In a particular embodiment the nucleic acid consists of 9 to 96 nucleotides. In another embodiment the nucleic acid consists of 12 to 48 nucleotides. In still another embodiment the nucleic acid consists of 15 to 36 nucleotides. In a preferred embodiment the nucleic acid consists of 17 to 20 nucleotides.

The present invention also provides a nucleic acid that hybridizes to the nucleotide sequence of SEQ ID NO:41, but not to the nucleotide sequence of SEQ ID NO:42 when the hybridization is performed under identical conditions. In a particular embodiment the nucleic acid comprises the nucleotide sequence of CCCACGGTCGGGTACCTGGGCGGGACGCGCCAGGCCACTCCCGGCGA (SEQ ID NO:29). The present invention further provides a nucleic acid that hybridizes to the nucleotide sequence of SEQ ID NO:42, but not to the nucleotide sequence of SEQ ID NO:41 when the hybridization is performed under identical conditions. In a particular embodiment the nucleic acid comprises the nucleotide sequence of CCCACGGTCGGGTGGCCGACTCCCGGCGA (SEQ ID NO:37).

In a related embodiment the present invention provides an isolated nucleic acid that hybridizes to the complementary strand of the nucleotide sequence of SEQ ID NO:42, but not to the complementary strand of the nucleotide sequence of SEQ ID NO:41 when the hybridization is performed under identical conditions. In still another embodiment the nucleic acid hybridizes to the nucleotide sequence of SEQ ID NO:41, but not to the nucleotide sequence of SEQ ID NO:42 when the hybridization is performed under identical conditions. In still another embodiment the nucleic acid hybridizes to the complementary strand of the nucleotide sequence of SEQ ID NO:41, but not to the complementary strand of the nucleotide sequence of SEQ ID NO:42 when the hybridization is performed under identical conditions.

The present invention also provides a nucleic acid that hybridizes to the nucleotide sequence of SEQ ID NO:42, but not to the nucleotide sequence of SEQ ID NO:45 when the hybridization is performed under identical conditions. In a related

embodiment the present invention provides an isolated nucleic acid that hybridizes to the complementary strand of the nucleotide sequence of SEQ ID NO:42, but not to the complementary strand of the nucleotide sequence of SEQ ID NO:45, when the hybridization is performed under identical conditions. In still another embodiment

5 the nucleic acid hybridizes to the nucleotide sequence of SEQ ID NO:45, but not to the nucleotide sequence of SEQ ID NO:42 when the hybridization is performed under identical conditions. In still another embodiment the nucleic acid hybridizes to the complementary strand of the nucleotide sequence of SEQ ID NO:45, but not to the complementary strand of the nucleotide sequence of SEQ ID NO:42 when the

10 hybridization is performed under identical conditions.

The present invention also provides for the use of the nucleic acids of the present invention (as well as other nucleic acids which can be used to identify DNA polymorphisms in the alleles of the genes involved in folate, pyridoxine, and/or cobalamin metabolism) in the methods of the present invention for identifying,

15 diagnosing, preventing and/or treating individuals.

In methods of estimating the susceptibility due to genetic or genetic and environmental factors for an individual to have or to develop a developmental disorder or to have offspring that develop a developmental disorder, and for the corresponding methods of generating genetic, or genetic and environmental reference

20 datasets, the present invention provides a step of analyzing nucleic acids and/or proteins from biological samples. In one particular embodiment, the assaying for the presence of the genetic variant of human dihydrofolate reductase having a nucleotide sequence with a 19 base-pair deletion spanning nucleotides 540 to 558 of the nucleotide sequence of SEQ ID NO:41 is included as part of this analysis. This

25 genetic variant of human dihydrofolate reductase becomes a genetic explanatory variable.

Determining if the biological sample contains the genetic variant of human dihydrofolate reductase having a nucleotide sequence with a 19 base-pair deletion spanning nucleotides 540 to 558 of the nucleotide sequence of SEQ ID NO:41 can be

performed by any appropriate method including PCR, special PCR, RT PCR, RFLP analysis, SSCP, and FISH.

In addition, all of the nucleic acids of the present invention including cDNA or genomic DNA can be placed into expression vectors operably associated with an expression control sequence. Alternatively, when the nucleic acid is part of an expression control sequence, the nucleic acid and/or the expression control sequence can be placed into an expression vector to control the expression of a coding sequence, such as a reporter gene. Such expression vectors can then be placed into either eukaryotic or prokaryotic host cells and expressed. The host cells comprising the expression vectors are also part of the present invention. In addition, when the nucleic acid includes a coding sequence or a part of a coding sequence, the present invention includes methods of purifying the gene products from the coding sequence or part thereof, and the purified gene products themselves.

Accordingly, it is a principal object of the present invention to provide a method for identifying an individual that is genetically inclined to develop a developmental disorder or disease.

It is a further object of the present invention to provide a method for identifying an individual that is genetically inclined to develop schizophrenia.

It is a further object of the present invention to provide a method for identifying an individual that is genetically inclined to have offspring having a developmental disorder.

It is a further object of the present invention to provide a method of diagnosing schizophrenia.

It is a further object of the present invention to provide a method of treating developmental disorders such as schizophrenia.

It is a further object of the present invention to provide a method for monitoring the treatment of the developmental disorder.

It is a further object of the present invention to provide a method for ameliorating the effect of a defect in folate, pyridoxine or cobalamin metabolism on a fetus due to the genetic or environmental status of a pregnant woman.

It is a further object of the present invention to provide a method of treating a patient who is genetically inclined to develop a developmental disorder such as schizophrenia.

It is a further object of the present invention to provide a method of overcoming a nutritional lack of folate, cobalamin or pyridoxine of a pregnant woman to prevent the development of the corresponding fetus developing a developmental disorder.

Other objects and advantages will become apparent to those skilled in the art from a review of the ensuing description.

These and other aspects of the present invention will be better appreciated by reference to the following drawings and Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows primers for PCR amplification of the dihydrofolate reductase (DHFR) deletion polymorphism region.

Figure 2 shows the genotypes of the DHFR 19 basepair deletion by non-denaturing polyacrylamide gel electrophoresis. Lanes 1 and 2 show genotypes 1,1. Lanes 3 and 4 show genotypes 1, 2. Lanes 5 and 6 show genotypes 2,2. Lane 7 shows phiX174 RF DNA/HaeIII size markers from BRL Life Technologies.

Figure 3 shows the sequences of PCR amplification products in the Region of the DHFR polymorphism region. * is explained in Text, *see* Example 2.

Figure 4A is a nucleotide sequence of the wild type human DHFR, (SEQ ID NO:41) from Yang *et al.*, *J. Mol. Biol.* **176**:169-187 (1984), GeneBank accession no: X00855.

The start codon is in bold. Figure 4B is the same nucleotide sequence as that of Figure 4A except the deletion of the 19 nucleotides due to the DHFR deletion

5 polymorphism, (SEQ ID NO:42).

DETAILED DESCRIPTION OF THE INVENTION

The present invention in its broadest embodiment provides a method of diagnosing, preventing and/or treating specific physiological/developmental disorders. Such physiological/developmental disorders include schizophrenia, spina bifida cystica,
10 Tourette's syndrome, bipolar illness, autism, conduct disorders, attention deficit disorder, obsessive compulsive disorder, chronic multiple tic syndrome and learning disorders such as dyslexia.

A particular aspect of the present invention provides methodology for diagnosing, preventing and/or treating a developmental disorder such as schizophrenia. Such
15 methodology is premised on the correlation between abnormalities in folate, cobalamin, and/or pyridoxine metabolism in an individual and/or the mother of an individual and the occurrence of the developmental disorder, *e.g.*, schizophrenia in the individual. Further, the present invention provides a framework (*i.e.*, the generatogen model, and the DNA Polymorphism-Diet-Cofactor-Development both of
20 which are described in detail below) which fully explain the rationale for the correlation, though the ultimate usefulness of the methods of the present invention are independent of any particular model.

Within this context, the DNA Polymorphism-Diet-Cofactor-Development model maintains that a developmental disorder such as schizophrenia results in part from
25 developmental brain damage sustained *in utero* due to maternal dietary deficiency of folate, pyridoxine or cobalamin potentiated by the aggregate effect of minor defects of folate, pyridoxine or cobalamin genes. The maternal damage to the fetus can result in part from insufficiency of the folate, pyridoxine and cobalamin themselves and/or from resulting effects such as immune deficiency and maternal teratogens, *e.g.*

hyperhomocysteinemia. Genes from either parent acting in the fetus may modify these damaging effects as exemplified in the gene-teratogen model, below.

As described herein the present invention can be practiced on a case by case basis, or alternatively, it can be used in the screening of the general population, or within any particular subgroup, such as newborns (as is presently performed in the diagnosis and treatment of hyperphenylalaninemia).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

As used herein a "gene involved in folate, pyridoxine, or cobalamin metabolism" is a gene that encodes a peptide or protein that plays a role in a pathway involved in either folate, pyridoxine, or cobalamin metabolism. An incomplete listing of examples of such proteins is given in Tables 2-7.

As used herein the term "individual" includes a fetus, infant, child, adolescent, and adult. Therefore, as used herein, an individual originates at conception.

As used herein an individual with a susceptibility for "having offspring that develop a developmental disorder" is meant to be indicative of the susceptibility of the offspring of that individual to develop the developmental disorder and is not in any way meant to be indicative of the susceptibility of the individual to have offspring.

The term "proband" as used herein is operationally defined by Table 8 along with the accompanying explanatory information (*see*, Example 1). For most purposes, the proband can be considered the central figure in the familial analysis, the remaining individuals in the family being designated as "blood relatives". There are three types of probands: (1) an "affected proband" *i.e.*, an individual that is believed to have a developmental disorder; (2) a "control proband" an individual that is believed not to have a developmental disorder; and (3) a "diagnostic proband" *i.e.*, an individual being diagnosed.

As used herein a "blood relative" of an individual is a relative that is related to the individual in a genetic sense. Blood relatives can include mothers, fathers, children, uncles, aunts, brothers, sisters, and grandparents. Preferably a blood relative is a parent, a sibling, or a grandparent. Adopted relatives, step-parents, relatives through marriage and the like are not blood relatives. Therefore, as used herein, the terms "mother", "father", "sibling", "grandparent", "grandfather" and "grandmother" are indicative of blood relationships.

As used herein a "mate of an individual" is a person whose genetic material is combined with that of the individual for the conception of the offspring in question.

- 10 As used herein the term "schizophrenia" describes a disorder that is at least partially due to one or more genetic mutations or polymorphisms in one or more genes involved in folate, cobalamin or pyridoxine metabolism in an individual that is schizophrenic and/or to one or more genetic mutations or polymorphisms in one or more genes involved in folate, cobalamin or pyridoxine metabolism in the mother of
15 that individual.

- As used herein an individual is "schizophrenic" when the individual displays symptoms that would be accepted by an experienced psychiatrist to merit a diagnosis of schizophrenia. Such a diagnosis is based, at least in part, on the currently evolving guidelines for the diagnosis of schizophrenia which are listed in the successive
20 editions of Diagnostic and Statistical Manual for Mental Disorders, put out by the American Psychiatric Association. The current edition is the DSM, Fourth Edition (1994).

- As used herein the terms "spina bifida cystica", "Tourette's syndrome", "bipolar illness", "autism", "conduct disorder", "attention deficit disorder", "obsessive
25 compulsive disorder", "chronic multiple tic syndrome" and "learning disorders" such as "dyslexia" describe disorders which display symptoms that would be accepted by an experienced psychiatrist to merit a diagnosis of that disorder. Such a diagnosis is based, at least in part, on the currently evolving guidelines which are listed in the successive editions of Diagnostic and Statistical Manual for Mental Disorders, put out

by the American Psychiatric Association. The current edition is the DSM, Fourth Edition (1994).

As used herein the term "teratogenic locus" indicates one or more alleles that act in a pregnant woman to cause an intrauterine teratogenic effect on the fetus.

- 5 As used herein the terms "specificity locus" or "modifying locus" are used interchangeably and are indicative of one or more alleles that can act during pregnancy and/or after birth to prevent, modify, and/or ameliorate the teratogenic effect of the teratogenic locus.

- 10 As used herein a "constellation of genetic mutations" is the set of genetic risk factor mutations that is present in a proband and relatives of the proband. One example of a constellation of genetic mutations is shown in a line of Table 8, below.

- 15 As used herein a "risk factor" is a teratogen or substance (including a defective gene) that can lead to a teratogenic effect that is present or suspected of being present in a tissue sample or body fluid of an individual's mother during the individual's gestation and/or present or suspected of being present in a tissue sample or body fluid of the individual.

- 20 As used herein a "genetic risk factor" is used interchangeably with the term "genetic explanatory variable" and is a genetic mutation and/or polymorphism that causes or potentially can cause the formation of and/or lead to the development of a risk factor in an individual or the individual's mother during gestation.

As used herein an "environmental risk factor" is used interchangeably with the term "environmental explanatory variable" and is an environmental factor that causes or potentially can cause the formation of and/or lead to the development of a risk factor in an individual or the individual's mother during gestation.

As used herein an “explanatory variable” is either an “environmental explanatory variable” or a “genetic explanatory variable” or the variable defined by their interaction or any combination of the above.

Enzymes whose deficiency may raise plasma homocysteine include

- 5 methylenetetrahydrofolate reductase (MTHFR), methionine synthase, and folate receptors/transport proteins/binding proteins (as well as all of the proteins listed in Tables 2-7 below).

The current (developmental) model for schizophrenia is that genetic and environmental factors cause brain damage in a fetus that later develops schizophrenia.

- 10 However, the genetic and environmental factors have not been identified. Also, extensive linkage and association studies have failed to identify genes determining schizophrenia. The reasons usually given for this difficulty include: (i) locus heterogeneity, *i.e.*, more than one gene locus is involved, perhaps many gene loci each with a small effect; (ii) the mode of inheritance of schizophrenia is unknown;
- 15 and (iii) an additional possible factor is that the frequency of the disease alleles may be high, thus greatly reducing the power of linkage studies.

The DNA Polymorphism-Diet-Cofactor-Development model explains all of these difficulties and at the same time proposes a unified metabolic abnormality. The unified metabolic abnormality is: (a) ENVIRONMENTAL, *i.e.*, due to a

- 20 folate/cobalamin/pyridoxine deficiency caused by either decreased ingestion or increased requirement during pregnancy; (b) GENETIC, *i.e.*, due to a folate/cobalamin/pyridoxine genetic defect caused by the aggregate effect of multiple mutations of folate/cobalamin/pyridoxine genes each individually having a small effect; and (c) the interaction of the folate/cobalamin/pyridoxine environmental and
- 25 genetic factors (indicated above) to cause other harmful effects such as maternal teratogens and immune deficiency during gestational development. Different gene loci and different combinations of gene loci will be involved in different patients and different families. The problem of locus heterogeneity is addressed by the hypothesis that the folate/cobalamin/pyridoxine genetic defect is the aggregate effect of multiple

mutations of folate/cobalamin/pyridoxine genes each of which have a relatively small effect.

The problem of mode of inheritance is addressed by the gene-teratogen model. The gene-teratogen model describes the special features of genes acting *in utero*; both

- 5 teratogenic and modifying of specificity loci may be involved. If these effects are not taken into account, the assignment of affection status in schizophrenia pedigrees is inaccurate. Assignment of affection status is a key element in defining the mode of inheritance for all kinds of linkage mapping. Failure to assign the correct mode of inheritance is another factor that has made the linkage studies very difficult.

- 10 Finally, the DNA Polymorphism-Diet-Cofactor-Development model proposes that some of the genetic factors for schizophrenia are common in the population. In fact, subclinical deficiency of folate, pyridoxine, and cobalamin is common in the population and common among pregnant women as well. Pregnancy further increases the requirement for folate, pyridoxine, and cobalamin. Common genetic
- 15 polymorphisms of folate and cobalamin genes are also known, some of them functional. Common genetic risk factors tend to be functional polymorphisms and/or mutant alleles that individually have small effects. Otherwise, they would be largely eliminated from the population by natural selection and would not be common. High disease allele frequency is yet another factor that greatly diminishes the power of a
- 20 linkage study.

Besides explaining the difficulties with current linkage studies, the DNA Polymorphism-Diet-Cofactor-Development model explains all of the unusual biological and epidemiological features of schizophrenia: *e.g.* the decreased amount of gray matter in brain areas, the unusual birth-month effect, the geographical

25 differences in incidence, the socioeconomic predilection, the association with obstetrical abnormalities (low birth weight and prematurity), and the association with famine and viral epidemics. Consistently, genetic linkage and cytogenetic studies in schizophrenia have implicated various chromosome regions, some of them containing folate, pyridoxine, and cobalamin genes including dihydrofolate reductase,

- 30 thymidylate synthase, and transcobalamin II. The DNA

Polymorphism-Diet-Cofactor-Development model predicts that folate, pyridoxine, or cobalamin gene mutations have a high frequency in schizophrenia patients or family members. Furthermore, mothers of schizophrenics are predicted to be particularly susceptible to producing one or more teratogens during pregnancy.

- 5 The present invention therefore provides methods for: (a) Diagnostic testing of schizophrenia by identifying a folate, pyridoxine, or cobalamin gene mutation or constellation of mutations in the patient, mother, and father. (b) Prevention of schizophrenia by diagnostic testing in families already affected by schizophrenia or by diagnostic population screening for folate mutations and identifying couples at risk
- 10 for producing schizophrenic offspring. These pregnancies can be further monitored for risk factors, *e.g.* dietary folate/pyridoxine/cobalamin, plasma folate/pyridoxine/cobalamin, or red blood cell folate; plasma homocysteine or other teratogens. (c) Therapy for schizophrenia, *e.g.*, treating the pregnant mother with folate, pyridoxine, cobalamin or other agents. The treatment can be monitored at
- 15 regular intervals to determine the effect of therapy. (d) Presymptomatic treatment of schizophrenia on young children found to be susceptible to schizophrenia by diagnostic testing for folate gene mutations and other risk factors can also be treated with methylfolate or related therapeutic modalities to forestall the appearance of schizophrenia symptoms in adolescence or adulthood.
- 20 Empirical studies with methylfolate treatment of schizophrenia have shown modest clinical improvement. The DNA Polymorphism-Diet-Cofactor-Development model gives a rationale for such therapy as well as for intensive testing of related therapeutic modalities. Genetic testing will need to be carried out in such patients to gauge their likelihood of responding to therapy. In addition, the DNA
- 25 Polymorphism-Diet-Cofactor-Development model gives direction and impetus toward uncovering the mechanism of fetal brain damage leading to schizophrenia.

Diagnostic testing for schizophrenia can involve testing not just the patient, but mother and father as well, for not just one factor but multiple genetic factors. For example, data for two gene loci (both folate-related genes) were used in Example 2.

- 30 In this case, there were only four explanatory variables for each comparison.

In addition, risk factors appearing only during pregnancy may play a role, *e.g.* dietary folate which can be further monitored during the pregnancy. In certain instances, genotype data can be used as the sole explanatory variables, particularly in the case when no environmental explanatory variables are known. In such a case, the

5 predicted probabilities will be only for the genetic component of the proband's risk of schizophrenia. In addition, schizophrenia mothers, fathers, and sibs do not necessarily have to come from the same families as the schizophrenia probands, as described in Example 2.

Of course certain genetic factors will turn out to be more common than others. This

10 may simplify testing somewhat. Also some genetic factors may operate chiefly in the mother, while others will operate chiefly in the schizophrenic patient. This may also simplify testing. There are some approaches to assessing risk factors during a past pregnancy, *e.g.* current dietary history as an indicator of past diet, methionine loading as in indicator of how susceptible a mother is to raising her plasma homocysteine,

15 assessment of other risk factors besides folate metabolism that may affect pregnancy outcome. Procedures including all of these variables are both envisioned and included in the present invention.

Thus the present invention provides a method of diagnosis of schizophrenia. In one aspect of the invention, diagnostic testing for genetic susceptibility to schizophrenia

20 determines the probability that the proband is affected with schizophrenia due to genetic factors. This is carried out by genetic testing of a patient suspected of having schizophrenia and/or whatever informative relatives are available, *e.g.* mother, father, sibs, or children. The genotypes of certain folate and/or cobalamin and/or pyridoxine gene mutations or constellation of mutations (folate and/or cobalamin and/or

25 pyridoxine gene mutations) are determined for each individual.

Since the abnormal phenotype of schizophrenia can be determined by both genetic and environmental factors and since other genetic factors besides folate/cobalamin/pyridoxine gene mutations may be involved, the presence of folate/cobalamin/pyridoxine gene mutations may be neither necessary nor sufficient

30 to cause schizophrenia. Thus, an unaffected individual may have the same genetic

risk factors as an affected individual but may lack sufficient environmental factors to cause the abnormal clinical disease. Also, an affected individual may lack folate/cobalamin/pyridoxine gene mutations but may have other related or non-related genetic risk factors that caused the schizophrenia.

- 5 Therefore folate/cobalamin/pyridoxine gene mutations are used as explanatory variables (genetic risk factors) to calculate the predicted probability that an individual has genetic susceptibility to schizophrenia due to these mutations. Genetic variation can be expected to account for approximately about half of the risk of developing schizophrenia since the concordance rate in identical twins has been estimated to be
- 10 about 50%. The other half of the risk results from environmental factors due to their different positions in the uterus and to differences in the blood supply. The use of environmental factors as additional explanatory variables enhances this probability calculation, although this environmental data is more difficult to gather. Together, using both genetic and environmental explanatory variables, the predicted probability
- 15 that an individual is schizophrenic may approach 1.0.

- One likely situation for the use of the present methodology is in the diagnosis of a patient that has developed a psychosis. In such a case, the clinician is likely to be interested in determining the probability that this individual has schizophrenia. The number of blood relatives (preferably first degree relatives) of the patient-to-be
- 20 diagnosed, both unaffected and affected, could then be determined. The number of these who would contribute a blood sample for analysis, for example, could then be ascertained. It is preferable that the patient-to-be-diagnosed also contributes a blood sample, however in certain situations, this may not be an option. The availability of dietary and epidemiological information for environmental explanatory variables,
 - 25 especially from the patient and the mother, can also ascertained. Of course all relevant legal and ethical rules should be followed regarding informed consent for the genetic testing.

- Biological samples such as tissue or fluid samples (*e.g.*, 7 ml of blood in an EDTA-containing vacutainer, *see* Example 2, below), and obtainable environmental
- 30 data from the patient and family members are then collected. DNA is extracted from

the sample and genotypes for alleles of folate and/or cobalamin and/or pyridoxine genes are determined. The methods for genotyping depend upon the specific genetic markers used as explanatory variables. The methods for allele determination for two genetic markers are discussed in the Examples below.

- 5 Data of the genetic and environmental explanatory variables for the patient-to-be-diagnosed (proband) and participating family members are added to a reference data set preferably consisting of well-defined schizophrenia probands and family members, and control probands, and family members for whom data is available for many explanatory variables. As an approximation the control probands themselves also can be used as the controls for each proband family member class as shown in Example 2, below. Thus, as an approximation the control probands can be used as controls for the affected probands; and/or separately for the mothers of affected probands; and/or separately for the fathers of affected probands, etc. Another example of a use of the control probands is in the evaluation and/or analysis of a particular diagnostic proband. In this case, the approximation is obtained by adding the diagnostic proband to the group of affected probands and control probands.

- A model is then created consisting of the explanatory variables actually available from specific patient-to-be diagnosed and family members participating in the testing. This new combined data set (reference data set and data from patient-to-be-diagnosed with participating family members) is analyzed by binary logistic regression (e.g., using a statistical software package such as the SAS System embodied in Example 1 below, though other programs may be used) for the model chosen giving the predicted probability that a proband is affected with schizophrenia for all of the probands including the patient-to-be-diagnosed.

- 25 In a particular embodiment the model is modified and the goodness of fit for the patient-to-be-diagnosed is checked. The predicted probability that the patient-to-be-diagnosed has schizophrenia is compared with a classification table generated from the model used to determine the likelihood of false positives and false negatives.

The predicted probability that the patient-to-be-diagnosed is affected with schizophrenia, with the likelihood of false positive or false negative result, can then be forwarded to the clinician.

The methods for determining an individual's risk for developing schizophrenia taught
5 by the present invention can be used in a variety of settings. For example, the present invention also provides a therapy for schizophrenia. Empirical studies with methylfolate treatment of schizophrenia have shown modest clinical improvement. The DNA Polymorphism-Diet-Cofactor-Development model provides a rationale for such therapy as well as for intensive testing of related therapeutic modalities, *e.g.*
10 other cofactors such as cobalamin or pyridoxine. In addition, the DNA Polymorphism-Diet-Cofactor-Development model gives direction and impetus toward uncovering the mechanism of fetal brain damage leading to schizophrenia. Of course such therapy also can be provided on a case by case basis in order to gauge the likelihood of the patient of responding to such therapy, with the methodology for
15 diagnosis of the present invention enabling the skilled practitioner to assess that likelihood.

In addition, the present invention provides a method of identifying individuals that are likely to be aided by presymptomatic treatment for schizophrenia. For example, young children found to have a high risk for susceptibility to schizophrenia by
20 diagnostic testing can be treated with methylfolate or related therapeutic modalities to forestall the appearance of schizophrenia symptoms in adolescence or adulthood. The present invention further provides methodology for diagnostic testing for specific families already affected by schizophrenia.

The present invention further provides methodology for population screening for
25 folate/cobalamin/pyridoxine mutations to help identify couples at risk for producing schizophrenic offspring. Subsequent or concurrent pregnancies can then be monitored for environmental risk factors, and treated with folate, cobalamin, pyridoxine or other agents and monitored at intervals for the effect of therapy. Such monitoring can include measuring levels of folate, cobalamin, pyridoxine or
30 homocysteine in a particular tissue and/or fluid sample, such as blood.

Since schizophrenia is a developmental disorder, it is likely that these same risk factors discussed here for schizophrenia could play a role in other developmental disorders including spina bifida cystica, Tourette's syndrome, learning disorders including dyslexia, conduct disorder, attention-deficit hyperactivity disorder, bipolar illness, autism, and obsessive-compulsive disorder. Interestingly, the mode of inheritance of these disorders, like that of schizophrenia, has been difficult to determine despite the fact that a genetic component to the etiology of each has been documented. Therefore, methodology analogous to that exemplified herein for schizophrenia can be readily adapted for diagnosing and/or treating other such developmental disorders.

Nucleic Acids

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (herein "Sambrook *et al.*, 1989"); *DNA Cloning: A Practical Approach*, Volumes I and II (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed. 1984); *Nucleic Acid Hybridization* [B.D. Hames & S.J. Higgins eds. (1985)]; *Transcription And Translation* [B.D. Hames & S.J. Higgins, eds. (1984)]; *Animal Cell Culture* [R.I. Freshney, ed. (1986)]; *Immobilized Cells And Enzymes* [IRL Press, (1986)]; B. Perbal, *A Practical Guide To Molecular Cloning* (1984); F.M. Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)].

A "nucleic acid molecule" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded

helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules including restriction fragments, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (*see* Sambrook *et al.*, *supra*). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. High stringency hybridization conditions correspond to 50% formamide, 5x or 6x SSC. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids, the GC percentage, and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (*see* Sambrook *et al.*, *supra*, 9.50-10.51). For hybridization with shorter nucleic acids, *i.e.*, oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (*see* Sambrook *et al.*, *supra*, 11.7-11.8). Preferably a minimum length for a hybridizable nucleic acid (*e.g.*, a nucleotide

probe or primer such as a PCR or RT-PCR primer) is at least about 12 nucleotides; preferably at least about 18 nucleotides; and more preferably the length is at least about 27 nucleotides; and most preferably at least about 36 nucleotides. Specific probes and primers that can be used to distinguish specific variants of the nucleic acids encoding the proteins involved in folate, pyridoxine, and/or cobalamin metabolism are also part of the present invention.

Such nucleotide probes and primers can be labeled or used to label complementary DNA (where appropriate) by any number of ways well known in the art including using a radioactive label, such as ^3H , ^{14}C , ^{32}P , or ^{35}S , a fluorescent label, a boron label [U.S. Patent No: 5,595,878, Issued January 21, 1997 and U.S. Patent No: 5,876,938, Issued March 2, 1999 which are incorporated by reference in their entireties], and enzymatic tags such as urease, alkaline phosphatase or peroxidase. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In a specific embodiment, the term "standard hybridization conditions" refers to a T_m of 55°C , and utilizes conditions as set forth above e.g., 5X SSC. In a preferred embodiment, the T_m is 60°C ; in a more preferred embodiment, the T_m is 65°C .

A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. If the coding sequence is intended for expression in a eukaryotic cell, a

polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

"Transcriptional and translational control sequences" are DNA regulatory sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "signal sequence" is included at the beginning of the coding sequence of a protein to direct the protein to a particular site/compartment in the cell such as the surface of a cell. This sequence encodes a signal peptide, N-terminal to the mature polypeptide, that directs the host cell to translocate the polypeptide. The term "translocation signal sequence" is used herein to refer to this sort of signal sequence. Translocation signal sequences can be found associated with a variety of proteins native to eukaryotes and prokaryotes, and are often functional in both types of organisms.

Identification of Genetic Mutations

A biological sample can be obtained from an individual and/or a blood relative of the individual, and from appropriate controls, using a sample from any body component

including tissue punches, body fluids, and hair, as long as the biological sample contains nucleic acids and/or proteins/peptides. Thus the DNA, mRNA, proteins or peptides of the biological sample can be used to identify mutations and/or variants in genes involved in folate, pyridoxine, or cobalamine metabolism. The present invention therefore includes methods of detecting and quantifying these nucleic acids and/or proteins/peptides that can be used to identify genetic risk factors.

In a particular embodiment the DNA is extractable. A particularly useful source of DNA is blood. For example, 2.5- 40 mls of blood can be collected in a vacutainer containing EDTA. The blood sample is placed on ice and then centrifuged to separate plasma, red cells, and buffy coat. The separated fractions are then frozen at -80°C.

The DNA can be isolated from the buffy coat by a number of procedures well known in the art including using a QIAmp column DNA extraction procedure or the QIAGEN Genomic-tip method. The isolated DNA can be digested with a series of restriction enzymes, for example, and then the digested products can be hybridized with one or more particular nucleic acid probes designed from a particular gene to identify the gene and preferably to test for particular genetic mutations.

Preferably the genomic DNA can be amplified by PCR using appropriate primer pairs such as the primer pairs for the MTHFR or DHFR genes which were used in the Example below. The PCR amplified product can be sequenced directly, or alternatively be digested with one or more appropriate restriction enzymes. The resulting digested products can be separated *e.g.*, by column chromatography, or preferably by polyacrylamide or agarose gel electrophoresis. The isolated digestion products can be compared *e.g.*, by previously determined restriction maps, and/or alternatively, the digestion products can be sequenced directly. Alternatively, as in the case of DHFR, genetic polymorphisms can be detected through the use of restriction enzymes.

Although a restriction map of a gene is sufficient for the employment of the methods disclosed herein, in preferred embodiments the nucleotide sequences of the genes used in the testing steps are known. To this end a large sampling of such sequences are provided in Tables 2-7. (These sequences may also be used in the design of
 5 restriction maps.) Thus, initially each gene whether used separately or used in a constellation of genes is characterized by the sequencing of the wild type gene, preferably including the coding regions, introns, control sequences, and other non-coding regions. In addition, mutations of such genes found in the general population can also be characterized. With the recent advances in the sequencing of the human
 10 genome the present invention contemplates that additional sequence information will become publicly available, particularly with regard to mutations in relevant introns, and control sequences etc. which are not available in cDNA libraries. Such sequence information is fully envisioned to be incorporated into the on-going compilations of relevant DNA sequence databases of the present invention, as well as for its parallel
 15 use in the general methodology described herein. Thus DNA or mRNA or cDNA made from the mRNA can be used to identify mutations and/or variants in genes involved in folate, pyridoxine, or cobalamine metabolism.

There are many methods currently known in the art to identify variant/mutant DNA, all of which may be used in the present invention (*see e.g.*, internet address
 20 <http://www.ich.bpmf.ac.uk/cmgs/mutdet.htm>). Such methods include but in no way are limited to direct sequencing, array sequencing, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Malditof) [Fitzgerald *et al.*, *Ann. Rev. Biophys. Biomol. Struct.* **24**:117-140 (1995)], Polymerase Chain Reaction "PCR", reverse-transcriptase Polymerase Chain Reaction "RT-PCR", RNAase
 25 protection assays, Array quantitation *e.g.*, as commercially provided by Affymetrix, Ligase Chain Reaction or Ligase Amplification Reaction (LCR or LAR), Self-Sustained Synthetic Reaction (3SR/NASBA), Restriction Fragment Length Polymorphism (RFLP), Cycling Probe Reaction (CPR), Single-Strand Conformation Polymorphism (SSCP), heteroduplex analysis, hybridization mismatch using

nucleases (*e.g.*, cleavase), Southern, Northern, Westerns, South Westerns, ASOs, Molecular beacons, footprinting, and Fluorescent *In Situ* Hybridization (FISH). Some of these methods are briefly described below.

- PCR is a method for increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. PCR can be used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves introducing a molar excess of two oligonucleotide primers which are complementary to their respective strands of the double-stranded target sequence to the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of denaturation, hybridization, and polymerase extension can be repeated in order to obtain relatively high concentrations of a segment of the desired target sequence. The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and, therefore, this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified." [Mullis (U.S. Patent No. 4,683,195) and Mullis et al. (U.S. Patent No. 4,683,202)]
- In Ligase Chain Reaction or Ligase Amplification Reaction (LCR or LAR) four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. [Barany, Proc. Natl. Acad. Sci., **88**:189 (1991); Barany, PCR

Methods and Applic., 1:5 (1991); and Wu and Wallace, *Genomics* 4:560 (1989)]
 LCR has also been used in combination with PCR to achieve enhanced detection of
 single-base changes. Segev, PCT Public. No. W09001069 A1 (1990).

- Self-Sustained Synthetic Reaction (3SR/NASBA) is a transcription-based *in vitro*
 5 amplification system [Guatelli *et al.*, *Proc. Natl. Acad. Sci.*, **87**:1874-1878, 7797
 (1990); Kwok *et al.*, *Proc. Natl. Acad. Sci.*, **86**:1173-1177] that can exponentially
 amplify RNA sequences at a uniform temperature. The amplified RNA can then be
 utilized for mutation detection (Fahy *et al.*, *PCR Meth. Appl.*, 1:25-33 (1991). In this
 method, an oligonucleotide primer is used to add a phage RNA polymerase promoter
 10 to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that
 includes a second primer, reverse transcriptase, RNase H, RNA polymerase and
 ribo- and deoxyribonucleoside triphosphates, the target sequence undergoes repeated
 rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the
 area of interest.
- 15 RFLP can be used to detect DNA polymorphisms arising from DNA sequence
 variation. This method consists of digesting DNA with one or more restriction
 endonucleases (*e.g.*, EcoRI) and analyzing the resulting fragments by means of
 Southern blots [Southern, E., *Methods in Enzymology*, **69**:152 (1980)], as further
 described by Botstein, *et al.*, *Am. J. Hum. Genet.*, **32**:314-331 (1980) and White, *et*
 20 *al.*, *Sci. Am.*, **258**:40-48 (1988). Since a DNA polymorphism may create or delete a
 restriction site, the length of the corresponding restriction fragment with any given
 restriction enzyme could change. Once a difference in a restriction fragment length is
 identified it can be used to readily distinguish a particular polymorphism from the
 wild type DNA. Mutations that affect the recognition sequence of the endonuclease
 25 will preclude enzymatic cleavage at that site, thereby altering the cleavage pattern of
 that DNA. DNAs are compared by looking for differences in restriction fragment
 lengths. A technique for detecting specific mutations in any segment of DNA is
 described in Wallace, *et al.*, [*Nucl. Acids Res.*, **9**:879-894 (1981)]. It involves

hybridizing the DNA to be analyzed (target DNA) with a complementary, labeled oligonucleotide probe. Due to the thermal instability of DNA duplexes containing even a single base pair mismatch, differential melting temperature can be used to distinguish target DNAs that are perfectly complementary to the probe from target
5 DNAs that differ by as little as a single nucleotide. In a related technique, described in Landegren, *et al.*, Science, 41:1077-1080 (1988), oligonucleotide probes are constructed in pairs such that their junction corresponds to the site on the DNA being analyzed for mutation. These oligonucleotides are then hybridized to the DNA being analyzed. Base pair mismatch between either oligonucleotide and the target DNA at
10 the junction location prevents the efficient joining of the two oligonucleotide probes by DNA ligase.

When a sufficient amount of a nucleic acid to be detected is available, there are advantages to detecting that sequence directly, instead of making more copies of that target, (*e.g.*, as in PCR and LCR). Most notably, a method that does not amplify the
15 signal exponentially is more amenable to quantitative analysis. Even if the signal is enhanced by attaching multiple dyes to a single oligonucleotide, the correlation between the final signal intensity and amount of target is direct. Such a system has an additional advantage that the products of the reaction will not themselves promote further reaction, so contamination of lab surfaces by the products is not as much of a
20 concern. Traditional methods of direct detection including Northern and Southern blotting and RNase protection assays usually require the use of radioactivity and are not amenable to automation. Recently devised techniques have sought to eliminate the use of radioactivity and/or improve the sensitivity in automatable formats.

One such example is the Cycling Probe Reaction (CPR) [Duck *et al.*, BioTech., 9:142
25 (1990)]. CPR, uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the

remainder of the probe from the target DNA and allowing another probe molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may be carried through sample preparation.

Single-Strand Conformation Polymorphism (SSCP) is based on the observation that single strands of nucleic acid can take on characteristic conformations in non-denaturing conditions, and these conformations influence electrophoretic mobility. [Hayashi, *PCR Meth. Appl.*, 1:34-38, (1991). The complementary strands assume sufficiently different structures that one strand may be resolved from the other. Changes in sequences within the fragment will also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations (Orita, *et al.*, *Genomics* 5:874-879, (1989). The SSCP process involves denaturing a DNA segment (*e.g.*, a PCR product) that is labeled on both strands, followed by slow electrophoretic separation on a non-denaturing polyacrylamide gel, so that intra-molecular interactions can form and not be disturbed during the run. This technique is extremely sensitive to variations in gel composition and temperature.

In Fluorescent In Situ Hybridization (FISH), specific probes are designed which can readily distinguish the wild-type gene from the variant/mutant gene. Such methodology allows the identification of a variant/mutant gene through *in situ* hybridization (U.S. Patent No. 5,028,525, Issued July 2, 1991; U.S. Patent No. 5,225,326, Issued July 6, 1993; and U.S. Patent No. 5,501,952, Issued March 26, 1996. FISH does not require the extraction of DNA. In addition, procedures for separating fetal blood cells from maternal blood cells are well known in the art allowing the fetus and the mother to be analyzed from the same body fluid sample (*see* U.S. Patent No: 5,629,147, Issued May 13, 1997).

Similarly, antibodies raised against specific mutations and/or variants in the gene products of the genes involved in folate, pyridoxine, or cobalamine metabolism can be used to identify specific polymorphisms. Alternatively, antibodies raised against the wild type proteins can be used to detect and/or quantify the amount of wild type protein present in a given biological sample. In the case in which cross-reacting protein isn't synthesized by the cells of an individual, or is synthesized in significantly lower amounts than those of control subjects, such determinations can be used to identify a genetic risk factor. In addition, these antibodies can be used in methods well known in the art relating to the localization and activity of the gene products, *e.g.*, for Western blotting, imaging the proteins *in situ*, measuring levels thereof in appropriate physiological samples, etc. using any of the detection techniques known in the art. Furthermore, such antibodies can be used in flow cytometry studies, in immunohistochemical staining, and in immunoprecipitation which serves to aid the determination of the level of expression of a protein in the cell or tissue.

In the particular instance when the gene product is an enzyme, *e.g.*, dihydrofolate reductase, the enzymatic activity of a biological sample can be indicative of the presence of a genetic risk factor. In a particular embodiment, a decrease in an enzyme activity that is associated with folate, pyridoxine, or cobalamine metabolism can be indicative of the presence of the genetic risk factor. Such assays can be performed on multiple samples such as on a microplate reader [Widemann *et al.*, Clin Chem. 45:223-228 (1999)].

MODEL 1

The Gene-Teratogen Model for the Inheritance Pattern of Certain Developmental Disorders

Introduction:

It has long been known, *e.g.* from extensive studies of exogenous teratogens in inbred mice [Finnell and Chernoff, *Gene-teratogen* interactions: an approach to

- understanding the metabolic basis of birth defects, In Pharmacokinetics in Teratogenesis, Vol. II:97-109 *Experimental Aspects In Vivo and In Vitro*, CRC Press, Inc, Boca Ratan, Fl. (1987)], that teratogens may be influenced by genetic factors. It is less well known that the same gene defect may cause different clinical disorders
- 5 depending upon whether the metabolic effect of the gene defect is exerted during gestation *in utero* or during postnatal life. However, the consequences of gene-teratogen interactions in human pedigrees have not been extensively explored, especially the consequences for the use of linkage mapping to identify an unknown gene acting *in utero* to cause a developmental disorder. A number of common human
- 10 developmental disorders have been shown to have a genetic component to their etiology. However, for certain developmental disorders, the mode of inheritance has been difficult to determine and linkage studies have met with unexpected difficulties or have achieved limited success. These developmental disorders include spina bifida cystica [Chatkupt, *Am J Med Genet*, **44**:508-512 (1992)], Tourette's syndrome &
- 15 related disorders, *e.g.* obsessive-compulsive disorder and chronic multiple tics syndrome [Pauls, *Adv Neurol*, **58**:151-157 (1992); McMahon *et al.*, *Adv Neurol*, **58**:159-165 (1992); Heutink *et al.*, *Am J Hum Genet*, **57**:465-473 (1995); Grice *et al.*, *Am J Hum Genet*, **59**:644-652 (1996)], learning disorders, including dyslexia [Lewis, *et al.*, *Behav Genet*, **23**:291-297 (1993); Pennington, *J Child Neurol 10 Suppl*, **1**:S69-
- 20 S77 (1995)], conduct disorder [Lombroso *et al.*, *J Am Acad Child Adolesc Psychiatry*, **33**:921-938 (1994)], attention-deficit hyperactivity disorder [Lombroso *et al.*, *J Am Acad Child Adolesc Psychiatry*, **33**:921-938 (1994)], bipolar illness [Baron, *Acta Psychiatr Scand*, **92**:81-86 (1995); Benjamin and Gershon, *Biol Psychiatry*, **40**:313-316 (1996); Risch and Botstein, *Nature Genet*, **12**:351-353 (1996); Jamison and
- 25 McInnis, *Nature Med*, **2**:521-522 (1996); Morell, *Science*, **272**:31-32 (1996)], schizophrenia [Owen, *Psychol Med*, **22**:289-293 (1992); Cloninger, *Am J Med Genet*, **54**:83-92 (1994); Lander and Kruglyak, *Nature Genet*, **11**:241-247 (1995); Baron, *Acta Psychiatr Scand*, **92**:81-86 (1995); Benjamin and Gershon, *Biol Psychiatry*, **40**:313-316 (1996); Baron, *Am J Med Genet*, **67**:121-123 (1996)], autism [Lombroso
- 30 *et al.*, *J Am Acad Child Adolesc Psychiatry*, **33**:921-938 (1994)], and

obsessive-compulsive disorder in adults [Lombroso *et al.*, *J Am Acad Child Adolesc Psychiatry*, **33**:921-938 (1994)]. A recent article [Moldin, *Nature Genet.* **17**:127-129 (1997)] has reviewed "The maddening hunt for madness genes."

The present model addresses the question of the mode of inheritance of certain developmental disorders and proposes the "gene-teratogen model." The model suggests that the mode of inheritance of genes acting prenatally may in some cases be fundamentally different from that of genes acting postnatally. Even the same gene acting prenatally may produce a different disorder from that gene acting postnatally. The inheritance pattern in the gene-teratogen model is simple, but from the perspective of the patient with the developmental disorder is neither dominant nor recessive. Some disorders regarded as multifactorial, polygenic, or oligogenic may have this mode of inheritance. In the gene-teratogen model, genetically determined teratogen production by the mother during pregnancy damages the fetus producing the abnormal phenotype of a developmental disorder. The model is illustrated with two types of loci, 1. a teratogenic locus acting in the mother, and 2. a modifying or specificity locus acting in the fetus. Damage by the teratogen is influenced also by environmental factors. The model is interesting because it is simple and because teratogenic loci will be difficult to locate by parametric or non-parametric linkage mapping techniques due to misspecification of the affection status of both mother and affected children. A study design is suggested for identifying teratogenic loci. An example of the gene-teratogen model is the major intrauterine effect seen in offspring of phenylketonuric mothers. Certain developmental disorders whose mode of inheritance has been difficult to determine or whose genetic factors have been difficult to locate are candidates for the gene-teratogen model, including spina bifida cystica, Tourette's syndrome, learning disorders including dyslexia, conduct disorder, attention-deficit hyperactivity disorder, bipolar illness, schizophrenia, autism, and obsessive-compulsive disorder.

The Gene Teratogen Model

The model is described in Table 1 using two kinds of loci: a "teratogenic" locus and a "modifying" or "specificity" locus. The gene-teratogen model requires a teratogenic locus. One or more modifying or specificity loci may or may not be present. Also, two types of phenotypes are defined: 1. the teratogen-induced phenotype; and 2. the teratogenic phenotype, *i.e.*, the phenotype of a mother that produces a teratogenic effect during pregnancy. The two phenotypes are different for the teratogenic locus but are identical for the modifying or specificity loci.

TABLE 1
DIAGRAM OF THE GENE-TERATOGEN MODEL

Grandparents:	Maternal Grandmother AabbCCdd	Maternal Grandfather AaBbCcdd	Paternal Grandmother AAbbCcDd	Paternal Grandfather AAbbCCdd
Parents:	Mother aaBbCcdd		Father AAbbCcDd	
Child:	Child (fetus) with developmental disorder AabbccDd			
locus A:	teratogenic locus, recessive, acting in the mother to cause intrauterine teratogenic damage to the fetus.			
locus B:	teratogenic locus, dominant, acting in the mother to cause intrauterine teratogenic damage to the fetus.			
locus C:	modifying or specificity locus, recessive, acting in the fetus.			
locus D:	modifying or specificity locus, dominant, acting in the fetus.			

The teratogenic locus may be dominant (locus A) or recessive (locus B). This locus acts in the mother during pregnancy to cause an intrauterine teratogenic effect in the fetus. The teratogenic effect may result from the production of an endogenous teratogen, from potentiation of an exogenous teratogen, from a metabolic deprivation or imbalance or from some other mechanism. Only one teratogenic locus is required;

both locus A and locus B are shown on the same diagram for simplicity. A specificity or modifying locus may be dominant (locus C) or recessive (locus D). Such a locus acts during pregnancy or after to modify the extent of the developmental damage done by the teratogenic locus or even to prevent or repair the damage. For example,

5 for a teratogen acting at a certain time in development, locus C or D may determine whether brain or kidney is damaged, which structures of the brain are damaged, or whether damage occurs at all.

1. *Locus A, recessive teratogenic locus, acting in the mother:* The child is the patient with the abnormal phenotype of a specific developmental disorder, while mother,
- 10 father, and grandparents do not have the abnormal phenotype of that disorder (Table 1). Locus A acts in the mother during pregnancy causing her to produce the teratogenic effect that damages the developing fetus leading to the developmental disorder either in the fetus or postnatally in the child or adult. Since this locus is recessive in action, the mother, a homozygote (aa) for the disease allele, is the genetic
- 15 "patient." Her abnormal phenotype, the "teratogenic phenotype", is the trait of producing the teratogenic effect during pregnancy. Her fetus, damaged by the teratogenic effect *in utero*, does develop the teratogen-induced phenotype. However, the fetus is only a heterozygote (Aa) at locus A and thus lacks both the abnormal homozygous genotype at locus A and the abnormal teratogenic phenotype; *e.g.*, if the
- 20 fetus is a daughter, she will not produce the teratogenic effect later during pregnancy. Thus, the fetus is affected with the developmental disorder but is not the genetic "patient." Locus A, acting through a teratogenic effect, cannot be the only etiological factor for the developmental disorder. If it were, then all pregnancies of an aa mother would have the teratogen-induced phenotype which is not the case. Environmental
- 25 and/or other genetic factors, are required. An aa father will have the abnormal genotype, but not the abnormal teratogenic phenotype because he could never become pregnant.

2. *Locus B, dominant teratogenic locus acting in the mother*: The situation is the same as for locus A except that locus B is dominant in action (Table 1). The mother has the abnormal genotype, Bb, and the abnormal teratogenic phenotype. The fetus has the teratogen-induced phenotype but in the instance shown (Table 1) has neither the abnormal genotype, the teratogenic phenotype, nor even a copy of the disease allele. The maternal grandfather shown (Table 1) has the abnormal genotype, Bb, but does not have the teratogenic phenotype because he could never become pregnant.

3. *Environmental effects*: The teratogenic effect is modified by environmental factors, e.g. maternal dietary factors, infection, or ingestion of teratogen. These environmental factors may interact with locus A or B or may act independently. From the perspective of the fetus later to develop the developmental disorder (teratogen-induced phenotype), intrauterine teratogenic is an environmental not a genetic effect.

4. *Modifying or Specificity Loci Acting in the Fetus, Loci C & D*: These loci may interact with the teratogenic locus or the environmental factors to increase or decrease their effect, or alternatively could act independently. Such genetic factors may be recessive (locus C) or dominant (locus D). Genotypes and phenotypes of locus C and D behave conventionally with respect to the developmental disorder. For locus C and D, the fetus is with the developmental disorder is now the genetic "patient". Maternal teratogenic *in utero* is an environmental effect. It is thus possible that the same gene locus could act in part as a teratogenic locus and in part as a modifying or specificity locus.

DISCUSSION

The Example of Phenylketonuria: An example of the gene-teratogen model is the major intrauterine effect in maternal phenylketonuria (PKU). Phenylketonuria itself is a recessive postnatal disorder. Untreated homozygous PKU mothers and fathers both have elevated blood phenylalanine (hyperphenylalaninemia). However,

- heterozygous offspring of untreated PKU mothers (but not fathers) have an abnormal phenotype.[Koch *et al.*, *Acta Paediatr Suppl*, **407**:111-119 (1994); Allen *et al.*, *Acta Paediatr Suppl*, **407**:83-85 (1994); Abadie *et al.*, *Archives Pediatr*, **3**:489-486 (1996)]. Thus the elevated blood phenylalanine or other metabolite(s) in the mother
- 5 acts as a teratogen for the fetus. Note that the fetus of an untreated phenylketonuric mother does not have the phenotype of PKU (the "teratogenic phenotype"), but has a different phenotype (the "teratogen-induced phenotype").
- Phenylketonurics [Menkes, *Textbook of Child Neurology*, Lea & Febiger, Philadelphia (1990)] are normal at birth and develop a progressive disorder
- 10 postnatally characterized by vomiting, eczema, seizures (infantile spasms with hypsarrhythmia on electroencephalography), and mental retardation. The fetus of an untreated phenylketonuric mother [Menkes, *Textbook of Child Neurology*, Lea & Febiger, Philadelphia (1990)] has a congenital non-progressive disorder of fetal origin characterized by microcephaly, abnormal facies, mental retardation, congenital
- 15 heart disease, and prenatal and postnatal growth retardation. The PKU phenotype is a postnatal degenerative disorder; the phenotype of the PKU intrauterine effect is a developmental disorder. The teratogenic effect is not dependent upon the fetal genotype, although the fetus is an obligate heterozygote since the mother is a homozygote for phenylketonuria and the father (usually) has the normal genotype.
- 20 Thus, in phenylketonuria, a mutation at the same gene locus causes two distinct disorders depending upon whether the period of abnormal gene action is prenatal or postnatal. A fetus with the abnormal homozygous genotype who is carried by a heterozygous mother is protected *in utero*, but develops PKU postnatally. A heterozygous fetus carried by a mother with the abnormal
- 25 homozygous genotype is damaged *in utero* when the mother's genotype predominates, but is protected from PKU postnatally by its own genotype.

An Example from Studies in Inbred Mice: Finnell and Chernoff [*Gene-teratogen interactions: an approach to understanding the metabolic basis of birth defects*, In *Pharmacokinetics in Teratogenesis*, Vol. II:97-109 *Experimental Aspects In Vivo and*

- In Vitro*, CRC Press, Inc, Boca Ratan, Fl. (1987)] have reviewed a group of elegant experiments in inbred mice documenting that differences in susceptibility to exogenous teratogens can be regarded as a genetic trait that is determined by susceptibility or liability genes of either the maternal or fetal genotype [Finnell and Chernoff, *Gene-teratogen* interactions: an approach to understanding the metabolic basis of birth defects, In *Pharmacokinetics in Teratogenesis*, Vol. II:97-109 *Experimental Aspects In Vivo and In Vitro*, CRC Press, Inc, Boca Ratan, Fl. (1987)]; Finnell *et al.*, *Am J. Med. Genet.* **70**:303-311 (1997); Bennett *et al.*, *Epilepsia* **38**:415-423 (1997)]. For example, sensitivity to acetazolamine-induced ectrodactyly is determined by the presence of three genes, and the fetus must be homozygous for the recessive allele at all three loci in order to express the malformation. However, the inbred mouse models used do not mirror the human situation in at least three respects. First, the human population is an outbred population compared to these inbred mouse models. Consequently, the relevant genotypes may be highly variable among members of different families. Second, the inbred mouse experiments address the question of exogenous rather than endogenous teratogens. Third, the inbred mouse studies rely upon known or candidate susceptibility loci, whereas in humans, the problem has been to locate and identify disease unknown loci largely by using linkage mapping techniques.
- 20 *Implications for Linkage Mapping:*
Teratogenic Locus (Locus A or B): The gene-teratogen model has major implications for linkage mapping done with either parametric or non-parametric methods. The problem for both methods is incorrect assignment of affection status. In the lod score method, a genetic model of the disease is constructed and an affection status is assigned to each member of the pedigree. If the genetic model specified is wrong, the linkage results may be falsely positive or falsely negative [Terwilliger and Ott, *Handbook of Human Genetic Linkage*, Johns Hopkins Univ. Pr., Baltimore (1994)].

In developmental disorders resulting from the gene-teratogen model, the phenotype assignment for lod score analysis will be incorrect. The patient with the developmental disorder will be assigned the affected phenotype, whereas the patient is actually affected only for the teratogen-induced phenotype, but is unaffected for the teratogenic phenotype. Likewise, the mother will be assigned the unaffected phenotype for linkage analysis. Actually, she is unaffected only for the teratogen-induced phenotype, but is affected for the teratogenic phenotype. Lod scores should increase when phenotype assignments have been corrected. However, apparently dominant inheritance may in fact turn out to be pseudodominant if the mutant allele is common in the population. For non-parametric analysis, a similar misassignment occurs. In the case of affected sib-pairs, the affected sibs will be assigned the affected phenotype. Actually, the sibs are affected only for the teratogen-induced phenotype, but are unaffected for the teratogenic phenotype. The mother will be assigned the unaffected or unknown phenotype. Actually, she is unaffected only for the teratogen-induced phenotype but is affected for the teratogenic phenotype. Thus, the "affected sib-pair" families are likely to turn out to contain only a single sporadic case, since the only individual in the kindred affected with the teratogenic phenotype will be the mother.

For the transmission/disequilibrium test (TDT) [Spielman *et al.*, *Am J Hum Genet*, **52**:506-516 (1993); Ewens and Spielman, *Am J Hum Genet*, **57**:455-464 (1995)] the patient with the developmental disorder will be assigned the affected phenotype. Actually, the patient will be affected only for the teratogen-induced phenotype but will be unaffected for the teratogenic phenotype. The mother will be assigned the unaffected or unknown phenotype. Actually, she is unaffected only for the teratogen-induced phenotype but is affected for the teratogenic phenotype. The expectation of TDT is that alleles of a linked locus will show distortion from random transmission from mother (or father) to the patient. Since the patient is unaffected for the teratogenic phenotype, no transmission distortion from mother (or father) to child will be observed. Transmission distortion for alleles of a teratogenic locus will in fact

occur from the mother's parents to the mother, the actual patient for the teratogenic phenotype. But this will not be looked for because the phenotypes have been wrongly assigned. In addition, grandparents of the patients with the developmental disorder have probably not had DNA collected. Therefore, for the TDT, negative results may

5 occur for disease alleles of a teratogenic locus because incorrect phenotype assignments will have been made. When correct phenotype assignments have been made, transmission distortion to the mother from her parents should be expected for disease alleles of a teratogenic locus. Analogous misassignments are made in allelic association and haplotype relative-risk analyses [Falk and Rubinstein, *Ann Hu, Genet*, 10 51:227-233 (1987); Terwilliger and Ott, *Hum Hered*, 42:337-346 (1992); Thomson, *Am J Hum Genet*, 57:487-498 (1995)].

Modifying or Specificity Loci (Locus C and/or D): Since these loci behave in a conventional fashion, the phenotype assignments will be correct. Consequently, genes identified by conventional parametric or non-parametric linkage studies are

15 likely to be modifying or specificity loci. An important question for linkage mapping is the relative contribution to the abnormal phenotype of the developmental disorder made by the teratogenic locus versus that of a modifying or specificity locus. If the effect of a teratogenic locus is small, then loci identified by conventional linkage studies will be specificity or modifying loci and the mode of inheritance will be

20 Mendelian or multifactorial. If a teratogenic locus makes a major contribution to phenotype, then linkage mapping studies will not give a consistent answer and the mode of inheritance will be difficult to determine.

The presence of a teratogenic locus may be suspected if the maternal contribution to phenotype is different from or greater than the paternal contribution. For example,

25 the mother's relatives of spina bifida infants more frequently have affected children than the father's relatives. Suggested explanations for this observation have been mitochondrial inheritance, maternal effect, or genomic imprinting [Chatkupt, *Am J Med Genet*, 44:508-512 (1992)]. The operation of a teratogenic locus is another

explanation and is itself a form of maternal effect. For a recessive teratogenic locus, the mother's sisters would be at greatest risk of having offspring with the teratogen-induced phenotype.

- Implications for Definition of Phenotype:* All the pregnancies of a mother with the
- 5 teratogenic phenotype are at risk for the developmental disorder, the teratogen-induced phenotype. Yet only a few of the fetuses will be affected by the developmental disorder because of the action of environmental factors and/or the modifying or specificity loci. The action of the environmental factors is fully quantitative: depending upon the amplitude of the environmental effect, a mild,
 - 10 moderate, or severe teratogen-induced phenotype may result. In addition, the environmental factor may act at different times in fetal development producing qualitatively different phenotypes. Thus, quantitatively or qualitatively different teratogen-induced phenotypes may result from pregnancies of the same mother with the teratogenic phenotype. In addition, the action of the modifying or specificity loci
 - 15 may produce quantitatively or qualitatively different phenotypes in offspring of the same couple. Such different phenotypes may be diagnostically classified as different disorders. This may complicate attempts at associating specific loci with a specific teratogen-induced phenotype. All of the teratogen-induced phenotypes resulting from pregnancies of a mother with the teratogenic phenotype modified only
 - 20 by environmental factors are genetically indistinguishable. However, such teratogen-induced phenotypes affected also by the various modifying or specificity loci segregating among the offspring of a single couple are only partially genetically related.

- Methods to Identify Teratogenic Loci:* One effective approach to finding a putative
- 25 teratogenic locus is to carry out non-parametric linkage studies of families consisting of a patient affected with the developmental disorder, the patient's two (unaffected) parents, and the patient's four (unaffected) grandparents (Table 1). In such a family, the mother is the genetic patient but the other family members are not. Now, the

mother's nuclear family (the mother and her parents) is compared with the father's nuclear family (the father and his parents). In a haplotype relative risk study, the disease allele(s) of the teratogenic locus will occur more frequently in the mother compared with other alleles of her parents; the disease allele(s) of the teratogenic

5 locus will not occur more frequently in the father compared with other alleles of his parents. In a transmission/disequilibrium test, transmission distortion will be seen for the disease allele(s) of a teratogenic locus in the mother's nuclear family but not in the father's nuclear family. In an allelic association study, the disease allele will occur more frequently in mothers, patients (with the developmental disorder), and patient's

10 sibs (both affected and unaffected) than in unrelated control individuals. Disease allele frequency in fathers will not be distinguishable from that in control individuals.

Certain developmental disorders with a genetic component to etiology, whose mode of inheritance has been difficult to determine or whose genetic factors have been difficult to locate, including those mentioned earlier, are candidates for the

15 gene-teratogen model.

MODEL 2:

The DNA Polymorphism-Diet-Cofactor-Development Hypothesis for Schizophrenia and Other Developmental Disorders

Folate metabolism is complex. At least 30 gene loci are involved in absorption, transport, and metabolism of folate, and these are regulated by additional gene loci.

20 Any of these is potentially a genetic risk factor for schizophrenia, although MTHFR and DHFR are particularly good candidates. Likewise, genes encoding proteins involved in the pathways of other vitamin-cofactors may be genetic risk factors.

Two cofactors that may be of particular potential importance are cobalamin and pyridoxine. Cobalamin is relevant because its metabolism is closely intertwined with

25 that of folate. For example, cobalamin is required for the activity of methionine

synthase (MTR), a folate-related enzyme. Decreased cobalamin can affect folate metabolism through the folate trap. Pyridoxine is relevant because the pyridoxine-dependent enzyme cystathionine beta-synthase (CBS), along with the cobalamin-dependent enzyme MTR and folate pathways including MTHFR and DHFR all participate in catabolism of homocysteine, an amino acid that is suspected of being a teratogen during pregnancy. Also, kynureninase, an important enzyme affecting niacin metabolism and serotonin synthesis is pyridoxine-dependent. Therefore, mutations of the genes encoding such proteins, especially common polymorphisms, could play a role in the cause of schizophrenia.

- 10 Since folate, cobalamin, and pyridoxine are all dietary constituents, the dietary content of these cofactors could lead to an "environmental" generation of a risk factor for schizophrenia. In addition genes encoding proteins involved in folate, cobalamin, and pyridoxine metabolism and catabolism could be genetic risk factors for schizophrenia. Thus, the cofactors and the proteins involved in pathways relevant
- 15 to these cofactors can potentially have either or both environmental and genetic effects on the susceptibility of an individual on schizophrenia.

Since the genetic aspect of schizophrenia differs so profoundly from other disorders which have been identified by linkage mapping techniques, it is clear that a new model for the genetic connection to schizophrenia is required. Therefore, the DNA

20 Polymorphism-Diet-Cofactor-Development (DDCD) hypothesis, is disclosed herein.

- The DDCD hypothesis is that interacting genetic and environmental factors affecting the metabolism of folate, cobalamin, or pyridoxine or all of these, play a role in the etiology of schizophrenia. The genetic effect results from the aggregate effect of multiple mutations that individually, for the most part, have small effects on folate-,
- 25 cobalamin- or pyridoxine-related genes, some of which will be common in the population, and can act *in utero*. Environmental factors include dietary folate and cobalamin and pyridoxine. If schizophrenia results from mild deficiency during fetal

development of dietary folate, cobalamin, or pyridoxine potentiated by mild genetic susceptibility mutations of genes related to these cofactors and by pregnancy, then this would be difficult to document by linkage mapping techniques. An example of interaction of genetic and environmental factors is that genetic factors are important for incorporating dietary folate; the enzyme dihydrofolate reductase is required for conversion of dietary folate to folinic acid thus allowing dietary folate to enter the body's metabolic pathways. Another example is that folate and cobalamin requirements increase during pregnancy; thus pregnancy could potentiate the effects of mild genetic defects of mother, fetus, or both. Deficiencies of a vitamin are often part of a broader dietary deficiency affecting multiple nutrients in addition to the vitamin being measured.

Locus Heterogeneity: The metabolic pathways of folate, cobalamin, and pyridoxine are complex and related to each other. Multiple gene loci code for the enzymes and transport proteins are required (Tables 2-7). Thus, a defect of folate, cobalamin, or pyridoxine metabolism could result from the aggregate effect of multiple mutations each of relatively small effect interacting with environmental factors. Different individuals might have different combinations of mutations. Such a metabolic defect would be difficult to detect by linkage mapping techniques because of locus heterogeneity.

Alternatively, even if one genetic defect were sufficient to make an individual more susceptible to having schizophrenic offspring, for example, because of the large number of potential genetic factors, and the corresponding importance of environmental factors, elucidation of such an individual genetic defect would still be difficult unless, of course, the genetic defect caused a major effect. The difficulty in elucidating an individual genetic defect is magnified when the genetic factor acts in the mother, and not in the schizophrenic patient.

High Disease Allele Frequency: Numerous mutational variants of folate and cobalamin genes are known. Some of these have functional significance and in addition are sufficiently common in a given population to be regarded as genetic polymorphisms. However, these common alleles are unlikely to have a major harmful effect by themselves, for if they did they would become uncommon in the population in the absence of selection effects, and would likely appear as Mendelian disorders. Thus, the folate, cobalamin, or pyridoxine disease alleles related to schizophrenia would appear to be more likely those of minor deleterious effect or those with harmful effect only in the presence of environmental deficiencies or pregnancy. Such disease genes of high population frequency will be difficult to detect by linkage mapping methods because high disease allele frequency decreases the power of linkage studies [Terwilliger and Ott, *Handbook of Human Genetic Linkage*, John Hopkins Univ. Press, Baltimore, (1994)].

Developmental Genes: Folate, cobalamin, and pyridoxine defects act prenatally as well as postnatally. Folate, cobalamin, and pyridoxine metabolism are crucial for DNA synthesis and cell division, which are of disproportionate importance during brain development. Some defects of folate, cobalamin, or pyridoxine metabolism elevate blood homocysteine, a toxic and potentially teratogenic substance. Genes acting in the mother to damage the developing fetus, *e.g. via* the gene-teratogen model (Model 1, above), have a mode of inheritance that is neither dominant nor recessive with respect to the fetus. Attempts to assign a mode of inheritance in this situation will be unsatisfactory because affection status would be incorrectly assigned. The mode of inheritance of a developmental disorder resulting from a teratogenic locus would be regarded as either multifactorial or unknown. This is the situation with schizophrenia whose mode of inheritance is unknown. Use of an incorrect genetic model decreases the power of a linkage studies [Terwilliger and Ott, *Handbook of Human Genetic Linkage*, John Hopkins Univ. Press, Baltimore, (1994)].

Genes of Folate Metabolism: Folate metabolism is extremely complex [Rosenblatt, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New York: McGraw-Hill, pp. 3111-3128 (1995); Mudd *et al.*, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New York: McGraw-Hill pp. 1279-1327 (1995)]. At least 30 gene loci (Table 2) have been identified as folate-related. These contribute to folate mediated 1-carbon transfer reactions, binding, transport and metabolism of folate, and other functions. A number of these have been cloned and localized to a chromosomal region (Table 3).

TABLE 2
FOLATE-RELATED GENES/ENZYMES/TRANSPORTERS^a

	Folate-Related Genes/Enzymes/Transporters ^a	SEQ ID NO:
	methylenetetrahydrofolate reductase, MTHFR, MIM 236250	1
5	methionine synthase (methyltetrahydrofolate:L-homocysteine S-methyltransferase), MTR, MIM 156570	2
	dihydrofolate reductase, DHFR, MIM 126060	3
	folylpolyglutamate synthase, FPGS, MIM 136510	4
10	folate receptor 1, folate receptor alpha (FOLR1, adult; FR-alpha), MIM 136430	5
	folate receptor 2, folate receptor beta (FOLR2, fetal; FR-beta), MIM 136425 (a.a.)	6
	folate receptor 2-like (FOLR2L, fetal-like), MIM-none	
	folate receptor gamma (FR-gamma), MIM 602469	7
15	serine hydroxymethyltransferase 1, SHMT1, MIM 182144	8
	methylenetetrahydrofolate dehydrogenase, methylenetetrahydrofolate cyclohydrolase, 10-formyltetrahydrofolate synthetase (trifunctional enzyme, MTHFD), MIM 172460	9
	serine hydroxymethyltransferase 2, SHMT2, MIM 138450	10
20	thymidylate synthase, TYMS, MIM 188350	11
	GAR (5-phosphoribosylglycineamide) transformylase, GART, MIM 138440	12
	reduced folate carrier-1, RFC1. Probably identical to micromolar membrane transport protein, intestinal folate carrier-1 (IFC1), and neutral folate transport protein. MIM 600424	13
25	cystathionine beta-synthase, CBS, MIM 236200	14
	AICAR (5-phosphoribosyl-5-aminoimidazole-4-carboxamide) transformylase	15
	glutamate formiminotransferase, MIM 229100	
	forminotetrahydrofolate cyclodeaminase	
	5, 10-methylenetetrahydrofolate synthetase	16
30	10-formyltetrahydrofolate dehydrogenase, Mim 600249	

	Folate-Related Genes/Enzymes/Transporters*	SEQ ID NO:
5	glycine cleavage pathway (SHMT plus three enzymes): MIM 238331 Gly-decarboxylase MIM 238300 H-Protein MIM 238330 T-Protein MIM 238310	17 18 19
	cblG (affects function of MTR), MIM 250940	
	methionine adenosyltransferase 1, MAT1A, (ATP:L-methionine S-adenosyltransferase), MIM 250850	20
10	pteroyl polyglutamate hydrolase ("conjugase"), form 1	
	pteroyl polyglutamate hydrolase ("conjugase"), form 2	
	NAD-dependent enzyme methylene tetrahydrofolate dehydrogenase cyclohydrolase (a.a.)	21
	methionine adenosyltransferase 2, MAT2A, MIM 601468	22
15	5-methyltetrahydrofolate- homocysteine methyltransferase reductase (MTRR) MIM 602568; #Variant in MTRR linked to cblE MIM 236270	23
	methyltransferases	
	S-adenosylmethionine decarboxylase, MIM 180980	24
	decarboxylated S-adenosylmethionine:putrescine propylaminotransferase or spermidine synthetase (a.a.)	25
20	S-adenosylhomocysteine hydrolase, , MIM 180960	26
	betaine-homocysteine methyltransferase dimethylthetin-homocysteine methyltransferase	27
	gamma-cystathionase (L-cystathionine cysteine-lyase (deaminating)), MIM 602888	28
25	folic acid transport protein, MIM 229050	
	DHFR (exon 6 and 3' flanking region)	30
	kynureninase	35
	human DHFR, exons 1 and 2 [Chen <i>et al.</i> , <i>J. Biol. Chem.</i> 259 :3933-3943 (1984)]	36
30	*listed with alternate names, abbreviations, and MIM numbers; #cblE is a phenotype for a particular group of disorders of folate/cobalamin metabolism. (a.a.) indicates the amino acid sequence	

TABLE 3
LOCALIZED GENE LOCI RELATED TO FOLATE METABOLISM

	Gene/enzyme/transport protein	Location	References
5	MTHFR	1p36.3	Goyette <i>et al.</i> , (1994); *, **
	MTR	1q43	Cook and Hamerton, (1979); Mellman <i>et al.</i> , (1979) **
	DHFR	5q11.2-13.2	Weiffenbach <i>et al.</i> , (1991) Gilliam <i>et al.</i> (1989b) *, **
	FPGS	9cen-q34	Jones and Kao (1984): Walter <i>et al.</i> (1992) *, **
	MAT	10q22	**
10	FR	11q13.3-q14.1 11q13.3-113.5	Lacey <i>et al.</i> (1989), Ragoussis <i>et al.</i> , (1992); Ratnum <i>et al.</i> (1989); Walter <i>et al.</i> (1992): * Ragoussis <i>et al.</i> , (1992), **
	SHMT2	12q12-q14 12q13	Garrow <i>et al.</i> , (1993); Law and Kao, (1979) * **
	MTHFD	14q24	Rozen <i>et al.</i> , (1989), Jones <i>et al.</i> (1981), *, **
	LCCL	16pter-qter	*, **
	SHMT1	17p11.2	Garrow <i>et al.</i> , (1993) *, **
15	TYMS	18p11.31-p11.22 18p11.32	* Hori <i>et al.</i> , (1990); Silverman <i>et al.</i> , (1993)
	SAHH	20cen-q13.1	*
	GART	21q22.1	McInnis <i>et al.</i> (1993) Schild <i>et al.</i> (1990) Avramopoulos <i>et al.</i> (1993) Goto <i>et al.</i> (1993) *, **
	RFC1	21q22.2-22.3	Moscow <i>et al.</i> , (1995)

Gene/enzyme/transport protein	Location	References
CBS	21q22.3	Munke <i>et al.</i> , (1988)
<p>notes: MTHFR=methylenetetrahydrofolate reductase. MTS=methionine synthase. DHFR=dihydrofolate reductase. FPGS=folylpolyglutamate synthase. MAT=methionine adenosyltransferase, (ATP:L-methionine S-adenosyltransferase). FR=folate receptor complex: FR-alpha=FOLR1=folate receptor 1, adult; FR-beta=FOLR2=folate receptor 2, fetal; FR-gamma; FOLR2L=folate receptor 2-like. SHMT2=serine hydroxymethyltransferase 2, mitochondrial. MTHFD=5, 10-methylenetetrahydrofolate dehydrogenase, 5, 10-methylenetetrahydrofolate cyclohydrolase, 10-formyltetrahydrofolate synthase (trifunctional enzyme). LCCL=gamma-cystathionase (L-cystathionine cysteine-lyase (deaminating). SHMT1=serine hydroxymethyltransferase 1, soluble. TYMS=thymidylate synthetase. SAHH, S-adenosylhomocysteine hydrolase. GART=phosphoribosylglycineamide formyltransferase. RFC1=reduced folate carrier-1 (possibly identical to IFC1, intestinal folate carrier-1). CBS=cystathionine beta-synthase. Location information from GOD (*), from MIM (**).</p> <p>Goyette <i>et al.</i>, <i>Nat. Gen.</i> 7:195-200 (1994) Cook and Hamerton, <i>Cytogenet Cell Genet.</i> 25:9-20 (1979) Mellman <i>et al.</i>, <i>Proc. Natl. Acad. Sci.</i> 76:405-409 (1979) Weiffenbach <i>et al.</i>, <i>Genomics</i> 10:173-185 (1991) 20 Gilliam <i>et al.</i>, <i>Genomics</i> 5:940-944 (1989b) Jones and Kao, <i>Cytogenet Cell Genet.</i> 37: 499 (1984) Walter <i>et al.</i>, <i>Ann. Hum. Genet.</i> 56:212 (1992) Lacey <i>et al.</i>, <i>Am.J. Med. Genet.</i> 60:172-173 (1989) Ragoussis <i>et al.</i>, <i>Genomics</i> 14:423-430 (1992) 25 Ratnum <i>et al.</i>, <i>Biochem.</i> 28:8249-8254 (1989) Garrow <i>et al.</i>, <i>J. Biol. Chem.</i> 268:11910-11916 (1993). Law and Kao, <i>Cytogenet Cell Genet.</i> 24: 102-114 (1979) Rozen <i>et al.</i>, <i>Ann. Hum. Genet.</i> 44:781-786 (1989) Jones <i>et al.</i>, <i>Somat. Cell Genet.</i> 7:399-409 (1981) 30 Hori <i>et al.</i>, <i>Hum. Genet</i> 85:576-580 (1990) Silverman <i>et al.</i>, <i>Genomics</i> 15:442-445 (1993) McInnis <i>et al.</i>, <i>Genomics</i> 16:562-571 (1993) Schild <i>et al.</i>, <i>Proc. Natl. Acad. Sci</i> 87:2916-2920 (1990) Avramopoulos <i>et al.</i>, <i>Genomics</i> 15:98-102 (1993) 35 Goto <i>et al.</i>, <i>Neuromusc Disord.</i> 3:157-160 (1993) Moscow <i>et al.</i>, <i>Cancer Res.</i> 55:3790-3794 (1995) Munke <i>et al.</i>, <i>Am J. Hum. Gen.</i> 42:550-559 (1988)</p>		

Genes of Cobalamin Metabolism: Cobalamin metabolism is also complex [Benton and Rosenberg, In: *The Metabolic and Molecular Bases of Inherited Disease,*

- 40 Disease, Scriver *et al.* (eds), New York: McGraw-Hill, 3129-3149 (1995)]. At least 15 gene loci (Table 4) have been identified as cobalamin-related. These contribute to

- the binding, transport, and metabolism of cobalamin, and its functions. A number of these have been cloned and localized to a chromosomal region (5). Cobalamin metabolism is closely intertwined with that of folate. For example, cobalamin is required for the activity of MTR, a folate-related enzyme. Decreased cobalamin can
- 5 affect folate metabolism through the folate trap [Rosenblatt, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New York: McGraw-Hill, pp. 3111-3128 (1995); Quadros *et al.*, *Biochem. Biophys. Res. Commun.*, **222**:149-154 (1996)].

TABLE 4

COBALAMIN-RELATED GENES/ENZYMES/TRANSPORTERS^a

	Cobalamin-Related Genes/Enzymes/Transporters ^a	SEQ ID NO:
5	(gastric) intrinsic factor, GIF, MIM-261000 (combined deficiency of GIF & R-binder, MIM 243320)	31
	intrinsic factor receptor, IFCR, MIM-261100	
	transcobalamin I, TCI (an R-protein, plasma), MIM 189905	32
	transcobalamin III, TCIII (an R-protein, plasma), MIM-none	
	other R-proteins (R-binders, cobalophylins, haptocorrins), MIM 193090	
10	transcobalamin II, TCII MIM 275350	33
	transcobalamin II receptor, TCII receptor, MIM-none	
	methylmalonyl Co-A mutase, MCM (MUT locus), MIM 251000	34
15	cblF, lysosomal cbl efflux, MIM 277380	
	cblC, cytosolic cbl metabolism, MIM 277400	
	cblD, cytosolic cbl metabolism, MIM 277410	
	cblA, mitochondrial cbl reduction, (AdoCbl synthesis only), MIM 251100	
	cblB, cob(I)alamin adenosyltransferase, (AdoCbl synthesis only), MIM 251110	
20	cblE, methyltransferase-associated cbl utilization, MIM 236270	
	cblG, methyltransferase-associated cbl utilization, MIM 250940	
^a listed with alternate names, abbreviations, and MIM numbers		

TABLE 5

LOCALIZED GENE LOCI RELATED TO COBALAMIN METABOLISM

Gene/enzyme/transport protein	Location	References
MCM (MUT locus)	6p21.2-p21.1	Qureshi <i>et al.</i> (1994) *
5 IF/GIF	11q12-q13	Hewit <i>et al.</i> (1991) *
TCI (an R-protein, plasma)	11q11-q12.3	Johnston <i>et al.</i> , (1992) Sigal <i>et al.</i> , (1987), *
TCII	22q11.2-q13 22q12/13 border	Li <i>et al.</i> , (1995)
10	notes: MCM=methylmalonyl Co-A mutase; IF/GIF=(gastric) intrinsic factor; TCI=transcobalamin I; TCII=transcobalamin II. Location information from GDB (*), from MIM (**).	
15	Qureshi <i>et al.</i> , <i>Crit. Rev. Oncol. Hematol.</i> 17 :133-151 (1994) Hewit <i>et al.</i> , <i>Genomics</i> 10 :432-440 (1991) Johnston <i>et al.</i> , <i>Genomics</i> 12 :459-464 (1992) Sigal <i>et al.</i> , <i>N. Engl. J. Med.</i> 317 :1330-1332 (1987) Li <i>et al.</i> , <i>Biochem. Biophys. Res. Comm.</i> 208 :756-764 (1995)	

- Genes of Pyridoxine Metabolism:* Pyridoxine metabolism is also complex with three dietary forms convertible to pyridoxal phosphate [Whyte *et al.*, *Hypophosphatasia*, In: The Metabolic and Molecular Bases of Inherited Disease, Scriver *et al.* (eds), New York: McGraw-Hill pp. 4095-4111 (1995)] and many pyridoxine-related and
- 20 pyridoxine-dependent enzymes including decarboxylases and all aminotranferases (Table 6). A number of pyridoxine-related enzymes have been cloned and localized to a chromosomal region (Table 7). Pyridoxine metabolism is related to folate metabolism, especially 1-carbon transfer reactions: both serine
- hydroxymethyltransferases and the P-protein (glycine decarboxylase) of the glycine
- 25 breakdown system are pyridoxine-dependent.

TABLE 6SOME PYRIDOXINE-RELATED GENES/ENZYMES/^a

	1.	cystathionine beta-synthase, CBS,	MIM 236200
	2.	gamma-cystathionase,	MIM 219500
5		(L-cystathionine cysteine-lyase, deaminating), LCCL	
	3.	glycine cleavage system (GCS): glycine decarboxylase (P-protein)	
	4.	serine hydroxymethyltransferase 1, SHMT1,	MIM 182144
	5.	serine hydroxymethyltransferase 2, SHMT2,	MIM 138450
10	6.	kynureninase	MIM 278600
	7.	all aminotransferases, (<i>e.g.</i> ornithine-gamma-aminotranferases, OAT,)	MIM 258870
	8.	decarboxylases,	MIM 266100
15		<i>e.g.</i> glutamic acid decarboxylases, GAD1, GAD2,	
	9.	pyridoxamine(pyridoxine)-5'-phosphate oxidase	MIM 603287

^alisted with alternate names, abbreviations, and MIM numbers.

TABLE 7

SOME LOCALIZED GENE LOCI RELATED TO PYRIDOXINE METABOLISM

	Gene/enzyme	Location	References
1.	GAD2	2q31,	Bu et al., 1992)
5	2. GCS P-protein	9p13	Hamosh et al.1995)
	3. GAD1	10p11.23	Bu et al.1992)
	4. OAT	10q26	**
	5. SHMT2	12q12-14	Garrow et al., 1993; Law and Kao, 1979
10	6. LCCL	16pter-qter	*, **
	7. SHMT1	17p11.2	Garrow et al.1993 * **
	8. CBS	21q22.3	Munke et al.1988
	9. PNPO (PPO)		Ngo et al. 1998

*listed with alternate names, abbreviations, and MIM numbers.

- 15 Location information from GDB (*), from MIM (**).

notes: GAD2=glutamic acid decarboxylase 2, 67 kDa. GCS=glycine cleaving system, P-protein=glycine decarboxylase subunit. GAD1=glutamic acid decarboxylase 1, 65 kDa. OAT=ornithine-gamma-aminotranferases. SHMT2=serine hydroxymethyltransferase 2, mitochondrial. LCCL=gamma-cystathionase
 20 (L-cystathionine cysteine-lyase (deaminating). SHMT1=serine hydroxymethyltransferase 1, soluble. CBS=cystathionine beta-synthase. PNPO=pyridoxamine(pyridoxine)-5'-phosphate oxidase

References:

- Bu et al., *Proc. Nat. Acad. Sci.*, **89**:2115 (1992).
- 25 Hamosh et al., In: "The Metabolic and Molecular Bases of Inherited Disease", Scriver et al. (eds), New York: McGraw-Hill pp.1337-1348 (1995).
- Garrow et al. *J. Biol. Chem.* **268**:11910-11916 (1993).
- Law and Kao, *Cytogenet Cell Genet*, **24**: 102-114 (1979).
- Munke et al. *Am J. Hum. Gen.* **42**:550-559 (1988).
- 30 Ngo et al. *Biochemistry* **37**:7741-7748 (1998).

- Relevance of Folate, Cobalamine, And Pyridoxine to Schizophrenia:* There is considerable evidence that schizophrenia results, at least in part, from damage to brain development *in utero* that becomes symptomatic in late adolescence or early adulthood. The etiology of schizophrenia has both genetic and environmental
- 5 components. Because folate, cobalamin, and pyridoxine are all ingested and metabolized, they could potentially be both environmental and genetic factors for schizophrenia. Folate, cobalamin, and pyridoxine are relevant to schizophrenia in important ways. First, all of them are required for cell division because of their role in nucleic acid synthesis [Rosenblatt, In: *The Metabolic and Molecular Bases of*
 - 10 *Inherited Disease*, Scriver *et al.* (eds) New York: McGraw-Hill, pp. 3111-3128 (1995); Benton and Rosenberg, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New York: McGraw-Hill, 3129-3149 (1995)]. The developmental brain insult implicated in schizophrenia [Akbarian *et al.*, *Arch. Gen. Psychiatry*, **50**:169-177 (1993); Akbarian *et al.*, *Arch. Gen. Psychiatry*, **50**:178-187
 - 15 (1993)] is an abnormality of neurogenesis and neuronal migration, which are midtrimester events requiring cell division. Thus folate, cobalamin, and pyridoxine deficiencies could result in the widespread decreased grey matter volume observed in schizophrenia.

- Individuals that become schizophrenic later in life are more likely to be born during
- 20 the winter and early spring [Boyd *et al.*, *Schizophr. Bull.*, **12**:173-186 (1986); Kendell and Adams, *Br. J. Psychiatry*, **158**:758-763 (1991); O'Callaghan *et al.*, *Br. J. Psychiatry*, **158**:764-769 (1991)]; this corresponds to midtrimester in late fall & winter. Many folate- and pyridoxine-containing foods, *e.g.* dark green leafy vegetables, are less readily available in late fall & winter in northern climates.
 - 25 Seasonality was found to be a major determinant of micronutrient status including folate status in a population of pregnant and lactating women in The Gambia where folate deficiency was widespread [Bates *et al.* *Eur. J. Clin. Nutr.* **48**:660-668 (1994)]. Dietary cobalamin comes from animal foods, *e.g.* meat, dairy products, and fish, and prolonged dietary insufficiency is required to produce cobalamin deficiency unless a

- person is a strict vegetarian or already has subclinical deficiency [Sanders and Reddy, *Am. J. Clin. Nutr.*, **59**:1176S-1181S (1994)]. In fact, a significant fraction of the population already has subclinical deficiency for folate [Lewis *et al.*, *Ann. NY Acad. Sci.*, **678**:360-362 (1993)] and for [Carmel *et al.*, *Arch. Intern. Med.*, **147**:1995-1996 (1987); Pennypacker *et al.*, *J. Am. Geriatr. Soc.*, **40**:1197-1204 (1992); Naurath *et al.*, *Lancet*, **346**:85-89 (1995); Allen *et al.*, *Am. J. Clin. Nutr.*, **62**:1013-1019 (1995); Black *et al.*, *J. Nutr.*, **124**:1179-1188 (1994)]. Also, the dietary folate requirement increases during pregnancy [Scholl *et al.*, *Am. J. clin. Nutr.*, **63**:520-525 (1996); McPartlin *et al.*, *Lancet*, **341**:148-149 (1993)] and most women become folate
- 10 deficient during late pregnancy [Giles, *J. Clin. Pathol.*, **19**:1-11 (1966)]. Cobalamin deficiency is also common during pregnancy [Gadowsky *et al.*, *J. Adolesc. Health*, **16**:465-474 (1995)] although subnormal levels of vitamin B12 during pregnancy must be interpreted with caution [Metz *et al.*, *Am. J. Hemetol.*, **48**:251-255 (1995)]. An increase in schizophrenia births has also been noticed after winter famine [Susser and
- 15 Lin, *Arch. Gen. Psychiatry*, **49**:983-988 (1992)]; Susser *et al.*, *Arch. Gen. Psychiatry*, **53**:25-31 (1996)], a time when severe dietary deficiency of both folate and cobalamin is more likely. A temporary increase in the incidence of neural tube defects was reported in Jamaica 11-18 months following Hurricane Gilbert and was found to be associated with decreased dietary folate [Duff and Cooper, *Am J. Pub. Health* **84**:473-
- 20 476 (1994)].

- Schizophrenia is also associated with obstetrical complications, e.g. low birth weight and prematurity [Lewis and Murray, *J. Psychiatr. Res.*, **21**:413-421 (1987)]. Low birthweight and prematurity have also been associated with dietary folate deficiency during pregnancy Scholl *et al.*, *Am. J. clin. Nutr.*, **63**:520-525 (1996).
- 25 Hyperhomocysteinemia is a risk factor for unexplained recurrent early pregnancy loss [Wouters *et al.*, *Fertil. Steril.*, **60**:820-825 (1993)] and for abruptio placentae [Goddijn-Wesel *et al.*, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, **66**:23-29 (1996)]. Hyperhomocysteinemia may be related to defects in folate-, cobalamin-, or pyridoxine-dependent reactions [Naurath *et al.*, *Lancet*, **346**:85-89 (1995)].

Interestingly, stillbirths and schizophrenia share a similar seasonality of birth excess [Torrey *et al.*, *Schizophr. Bull.*, **19**:557-562 (1993)]. Also N_2O , an anaesthetic gas that inhibits MTR, a cobalamin-requiring enzyme of folate metabolism, is a reproductive toxin for both men and women [Louis-Ferdinand, *Adverse Drug React. Toxicol Rev.*, **13**:193-206 (1994)]. Methotrexate, an inhibitor of dihydrofolate reductase (DHFR), induces abortion.

- Dietary folate deficiency and low plasma folate are common in inner city urban populations [Scholl *et al.*, *Am. J. clin. Nutr.*, **63**:520-525 (1996)]. Likewise, schizophrenia has been reported to be more common in inner city urban populations
- 10 [Fuller and Bowler, *Schizophr. Bull.*, **16**:591-604 (1990)]. Also, both low folate intake [Schorah and Wild, *Lancet.*, **341**:1417 (1993)] and schizophrenia [Dohrenwned *et al.*, *Science*, **255**:946-952 (1992)] are correlated with lower socioeconomic status.

- Immune function is impaired in folate deficiency [LeLeiko and Chao, In: *Rudolph's Pediatrics*, 20th ed., Stamford, CT: Appleton & Lange, pp. 1001-1010 (1996)], in
- 15 cobalamin deficiency [Hitzig *et al.*, *Ciba. Found. Symp.*, **68**:77-91 (1978)] and in pyridoxine deficiency [Trakatellis *et al. Postgrad Med. J.* **73**:617-622 (1997)] and deficient individuals are more susceptible to infection. Methotrexate, an inhibitor of dihydrofolate reductase, inhibits immune function [Hughes, In: *Rudolph's Pediatrics*, 20th ed., Stamford, CT: Appletone and Lange, pp. 517-519 (1997)]. And, as
- 20 mentioned, dietary folate and cobalamin requirements increase during pregnancy [Scholl *et al.*, *Am. J. clin. Nutr.*, **63**:520-525 (1996); McPartlin *et al.*, *Lancet.*, **341**:148-149 (1993)]. This is relevant because the season-of-birth effect just mentioned in connection with dietary folate, or cobalamin deficiency has also been explained by *in utero* infectious illness, the "viral theory" of schizophrenia.
- 25 Individuals born following winters with severe influenza epidemics are more likely to develop schizophrenia [Adams *et al.*, *Br. J. Psychiatry*, **163**:522-534 (1993)] though not all studies find this effect. Although it has not been demonstrated that either the schizophrenia fetus or the pregnant mother actually developed influenza, the

histologic pattern in schizophrenia of a neuronal migration abnormality during brain development has been seen as compatible with a fetal viral infection [Kovelman and Scheibel, *Biol. Psychiatry*, **19**:1601-1621 (1984); Bogerts *et al.*, *Arch. Gen. Psychiatry*, **42**:784-791 (1985); Akbarian *et al.*, *Arch. Gen. Psychiatry*, **50**:169-177 (1993); Akbarian *et al.*, *Arch. Gen. Psychiatry*, **50**:178-187 (1993)]. Thus folate or cobalamin, deficiency during pregnancy could result in greater susceptibility to viral infection affecting mother, fetus, or both. The infectious agent could be influenza itself. Alternatively, a severe influenza epidemic could be a "marker" of a severe winter, and infection by another agent could cause the brain damage. In this way, folate or cobalamin deficiency could cause the season-of-birth effect either through the mechanism of dietary deficiency alone, through maternal immune deficiency and infection, or both.

Methotrexate, a DHFR inhibitor, is also an important therapeutic agent for rheumatoid arthritis. Rheumatoid arthritis has repeatedly been found to have a decreased frequency in schizophrenics, a puzzling finding that remains unexplained [Eaton *et al.*, *Schizophr. Res.*, **6**:181-192 (1992)].

The developmental model of schizophrenia postulates that brain damage sustained in the second trimester of fetal life results in schizophrenia later in development [Brixey *et al.*, *J. Clin. Psychol.*, **49**:447-456 (1993)]. Both folate and cobalamin are already known to contribute to a first trimester fetal nervous system malformation, spina bifida cystica [Kirke *et al.*, *Q. J. Med.*, **86**:703-708 (1993); Gordon, *Brain Dev.*, **17**:307-311 (1995)], and possibly other birth defects [Shaw *et al.*, *Lancet.*, **346**:393-396 (1995); Czeizel, *Lancet.*, **345**:932 (1995)]. Some studies [Whitehead *et al.*, *Q. J. Med.*, **88**:763-766 (1995); van der Put *et al.*, *Lancet.*, **346**:1070-1071 (1995); Ou *et al.*, *Am. J. Med. Genet.*, **63**:610-614 (1996); Chatkupt *et al.*, *Am. Acad. Neurol. Works in Progres*, **WIP4**: (1996)] suggest that a genetic susceptibility factor for spina bifida is a common allele of the folate gene, MTHFR, the nucleotide 677C->T transition converting an alanine residue to valine resulting in a heat-labile enzyme protein.

- Homozygotes for this allele, about 10% of the normal population, have lower erythrocyte folate and plasma folate during pregnancy [Molloy *et al.*, *Lancet.*, **349**:1591-1593 (1997)]. Homozygotes for this allele also develop moderately elevated blood homocysteine [van der Put *et al.*, *Lancet.*, **346**:1070-1071 (1995);
- 5 Frosst *et al.*, *Nature Genet.*, **10**:111-113 (1995)] in the presence of dietary folate deficiency. Moderate hyperhomocysteinemia is toxic to adults [Fermo *et al.*, *Ann. Intern. Med.*, **123**:747-753 (1995)], and toxic to the fetus in early gestation [Wouters *et al.*, *Fertil. Steril.*, **60**:820-825 (1993)], and possibly teratogenic in the first trimester causing neural tube defects [Whitehead *et al.*, *Q. J. Med.*, **88**:763-766 (1995); van der
- 10 Put *et al.*, *Lancet.*, **346**:1070-1071 (1995); Ou *et al.*, *Am. J. Med. Genet.*, **63**:610-614 (1996). Thus, the MTHFR heat-labile mutation, in the presence of decreased dietary folate in midtrimester, could be teratogenic both through hyperhomocysteinemia and also through folate deficiency causing the developmental brain damage hypothesized in the developmental model of schizophrenia [Brixey *et al.*, *J. Clin. Psychol.*, **49**:447-
- 15 456 (1993)]. A second common polymorphism of MTHFR, the nt1298 A->C mutation could also be a genetic risk factor for spina bifida [van der Put *et al.*, *Lancet.*, **346**:1070-1071 (1995)].

- Schizophrenia is a common disorder, affecting 1% or more of the population [Karno *et al.*, In: *Comprehensive Textbook of Psychiatry*/VI, 6th ed., Baltimore: Williams &
- 20 Wilkins, pp. 902-910 (1995)]. Thus, if a significant proportion of schizophrenia shares a common etiology, both the genetic susceptibility factors and the environmental factors must be common in the population. As mentioned earlier, a significant fraction of the population is already sub-clinically deficient for folate and for cobalamin; also, pregnancy may increase this fraction since dietary folate and
- 25 cobalamin requirements increase during that time. Several functional polymorphic alleles of folate and cobalamin genes are also common in the population including the MTHFR mutations just mentioned and polymorphisms of thymidylate synthase [Horie *et al.*, *Cell Struct. Funct.*, **20**:191-197 (1995)], transcobalamin II [Li *et al.*, *Biochim. Biophys. Acta.*, **1219**:515-520 (1994)], and folate-binding proteins [Li *et al.*,

improved symptoms of depression in an open trial in elderly depressed patients [Guaraldi *et al. Ann. Clin. Psychiatry* **5**:101-105 (1993)]. Schizophrenics are reported to have an 80% excess mortality from cardiovascular disease [Gottesman, *Schizophrenia Genesis*, Schizophrenia Genesis- The Origins of Madness, W.H.

- 5 Freeman & Co. N.Y.(1991)]; hyperhomocysteinemia, dietary folate deficiency and the MTHFR 677C->T mutation have been implicated in cardiovascular disease in some studies [Morita *et al., Circulation*, **95**:2032-2036 (1997)] but not others (Anderson *et al., J. Am. Coll. Cardiol.* **30**:1206-1211 (1997)]. Also, kynureninase, an important enzyme of tryptophan metabolism, affecting niacin metabolism and
- 10 serotonin synthesis, is pyridoxine-dependent. Niacin deficiency (pellagra) can cause mental changes including psychosis and hallucinations [Wilson, *Vitamin deficiency and excess*, pp.472-480. In: *Harrison's Principles of Internal Medicine*, (Scriber *et al.* e's.) McGraw-Hill, Inc., N.Y. (1994)]. Also, clozapine, resperidone, and olanzapine are thought to exert their antipsychotic effect in schizophrenia in part through
- 15 serotonin receptor antagonism.

Gene Localization Studies in Schizophrenia and Folate/Cobalamine/Pyridoxine

Genes: If folate, cobalamin, or pyridoxine genes are susceptibility factors for schizophrenia, it is possible that gene localization studies have already identified candidate chromosome regions that contain such a gene (Tables 3, 5, and 7). For three

- 20 folate or cobalamin genes, DHFR, TCNII and TYMS, there is excellent concordance with schizophrenia gene localization studies.

On chromosome 5, DHFR has been located at 5q11.2-13.2. A schizophrenia translocation [t(1;5)(1q32.3;5q11.2-13.3)] was reported [McGillivray *et al., Am. J. Med. Genet.*, **35**:10-13 (1990); Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)]

- 25 affecting 5q11.2-5q13.3. A proband and uncle, both with schizophrenia and eye-tracking abnormalities, had partial trisomy for 5q11.2-5q13.3; the third copy was inserted at 1q32.3 giving a derivative chromosome, der(1)inv ins(1;5)(q32.2;q13.3q11.2). The proband's mother had a balanced translocation but

was phenotypically normal without schizophrenia or eye-tracking abnormalities. She had the derivative chromosome 1 with extra material from chromosome 5 inserted but a corresponding deletion in one of her chromosomes 5. She thus had only two copies of 5q11.2-5q13.3. Further studies [Gilliam *et al.*, *Genomics*, **5**:940-944 (1989)]

- 5 showed that the DHFR gene is located within this deleted region, 5q11.2-13.3. Another schizophrenia chromosome abnormality, inv5(p13;q13), has been reported [Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)] affecting 5q13.

- On chromosome 5, two-point lod scores of 4.64 and 2.29 were found [Sherrington *et al.*, *Nature*, **336**:164-167 (1988)] for the polymorphic markers D5S76 and D5S39
10 respectively in the region of the chromosome abnormality just discussed [McGillivray *et al.*, *Am. J. Med. Genet.*, **35**:10-13 (1990); Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)] affecting 5q11.2-13.3. Two other linkage studies found small positive lod scores in this region [Coon *et al.*, *Biol. Psychiatry*, **34**:277-289 (1993); Kendler and Diehl, *Schizophr. Bull.*, **19**:261-285 (1993)], but numerous other studies excluded this
15 region under the assumptions and models used [Kendler and Diehl, *Schizophr. Bull.*, **19**:261-285 (1993)].

- On chromosome 18, TYMS has been located at 18p11.32-p11.22. A ring chromosome with deletion of 18pter-p11,18q23-qter [Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)] was reported in a kindred with schizophrenia and bipolar illness
20 [Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)]. Deletion of a segment of 18p was reported in a schizophrenia chromosome [Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)].

- On chromosome 22, TCNII has been located at 22q11.2-q13, possibly at the 22q12/13 border. High lod scores have consistently been obtained in the region of TCNII:
25 IL2RB, in 22q12-q13.1 gave a lod score [Pulver *et al.*, *Am. J. Med. Genet.*, **54**:3-43 (1994)] of 2.82. Other markers over a broad region of 22q have given suggestive lod scores. D22S278, in 22q12, gave a lod score [Vallada *et al.*, *Am. J. Med. Genet.*,

60:139-146 (1995)] of 1.51. CRYB2, in 22q11.2-q12.1, gave a lod score [Lasseter *et al.*, *Am. J. Med. Genet.*, **60**:172-173 (1995)] of 1.71. D22S10, in 22q11.1-q11.2, gave a lod score [Coon *et al.*, *Biol. Psychiatry*, **34**:277-289 (1993)] of 0.79. Highly significant p-values for non-parametric analyses have also been obtained: D22S278, 5 in 22q12, for example gave $p=.001$ [Gill *et al.*, *Am. J. Med. Genet.*, **67**:40-45 (1996)].

The deletions of velocardiofacial (VCF) syndrome and related disorders (DiGeorge syndrome (DGS) and CATCH22) are located [Lindsay *et al.*, *Genomics*, **32**:104-112 (1996)] at 22q11.2. A psychotic disorder develops in about 10% of patients with VCF syndrome [Chow *et al.*, *Am. J. Med. Genet.*, **54**:107-112 (1994)]. TCNII is not 10 known to be located at or within these deletions. VCF and related disorders are relatively uncommon compared to schizophrenia; only 2 of 100 randomly selected patients (92 schizophrenics, 5 with schizoaffective disorder, and 3 with schizophreniform disorder) in the Maryland Epidemiological Sample were found [Lindsay *et al.*, *Am. J. Hum. Genet.*, **56**:1502-1503 (1995)] to have VCF-related 15 deletions (and later VCF syndrome) on 22q11.2. Consequently, it is not clear whether schizophrenia linkage studies are detecting a haplotype related to a VCS locus or some other locus in this region, such as TCNII.

For some other folate, cobalamin, or pyridoxine relevant genes, physical or genetic studies of schizophrenia have identified chromosomal regions near the gene.

20

DISCUSSION

The folate-cobalamin hypothesis for schizophrenia is attractive because it suggests that a single mechanism of genetic and environmental factors may play a major role in the etiology and pathogenesis of schizophrenia. The combined result of this mechanism is to damage fetal development, especially brain development by 25 inhibiting nucleic acid synthesis, by affecting gene methylations, by increasing susceptibility to infection, and/or by producing teratogens.

This mechanism addresses several puzzling features of schizophrenia such as the season of birth effect, the association with famine and influenza epidemics, the negative association with rheumatoid arthritis, the associations with obstetrical abnormalities, social class, and urban environment. The mechanism also suggests approaches to diagnostic testing, to prevention, and to improved therapy.

It is not excluded that such a mechanism could also apply to a number of common human developmental disorders that have been shown to have a genetic component to their etiology but whose mode of inheritance has been difficult to determine and for which linkage studies have met with unexpected difficulties or have achieved limited success. These developmental disorders include Tourette's syndrome & related disorders (e.g. obsessive-compulsive disorder and chronic multiple tics syndrome) [Pauls, *Adv Neurol*, **58**:151-157 (1992); McMahon *et al.*, *Adv Neurol*, **58**:159-165 (1992); Heutink *et al.*, *Am J Hum Genet*, **57**:465-473 (1995); Grice *et al.*, *Am J Hum Genet*, **59**:644-652 (1996)], learning disorders, including dyslexia [Lewis, *et al.*, *Behav Genet*, **23**:291-297 (1993); Pennington, *J Child Neurol* **10 Suppl**, **1**:S69-S77 (1995)], conduct disorder [Lombroso *et al.*, *J. Am. Acad. Child Adolesc. Psychiatry*, **33**:921-938 (1994)], attention-deficit hyperactivity disorder [Lombroso *et al.*, 1994, *J. Am. Acad. Child Adolesc. Psychiatry*, **33**:921-938 (1994)], bipolar illness [Baron, *Acta. Psychiatr. Scand.*, **92**:81-86 (1995); Benjamin and Gershon, *Biol. Psychiatry*, **40**:313-316 (1996); Risch and Botstein, *Nature Genet.*, **12**:351-353 (1996); Jamison and McInnis, *Nature Med.*, **2**:521-522 (1996); Morell, *Science*, **272**:31-32 (1996)], autism [Lombroso *et al.*, 1994, *J. Am. Acad. Child Adolesc. Psychiatry*, **33**:921-938 (1994)], and obsessive-compulsive disorder in adults [Lombroso *et al.*, 1994, *J. Am. Acad. Child Adolesc. Psychiatry*, **33**:921-938 (1994)]. Some of these disorders have been shown to be associated with schizophrenia.

The present invention may be better understood by reference to the following non-limiting Examples, which are provided as exemplary of the invention. The following Examples are presented in order to more fully illustrate one embodiment of the

invention. They should in no way be construed, however, as limiting the broad scope of the invention.

EXAMPLE 1

DIAGNOSING SCHIZOPHRENIA

5

Structure of Datafiles

Data are arranged in a file suitable for input into a binary logistic regression program (Table 8). A model is created consisting of those explanatory variables actually available from the specific patient-to-be-diagnosed and family members participating in the testing. This new combined data set (reference data set + data from

10 patient-to-be-diagnosed with participating family members) is analyzed by binary logistic regression for the model chosen giving the predicted probability that a proband is affected with schizophrenia for all of the probands including the patient-to-be-diagnosed.

The model can be modified if required. The goodness of fit for the

15 patient-to-be-diagnosed is checked. The predicted probability that the patient-to-be-diagnosed has schizophrenia is compared with a classification table generated from the model used to determine likelihood of false positives and false negatives. The predicted probability that the patient-to-be-diagnosed is affected with schizophrenia, with likelihood of false positive or false negative result, is returned to

20 the clinician.

TABLE 8

A HYPOTHETICAL PARTIAL REFERENCE DATA SET OF GENETIC
EXPLANATORY VARIABLES TO ILLUSTRATE DATA STRUCTURE

ID	resp	P111	P112	P211	P212	M111	M112	M311	F511	S2-411	CA1-111
1	1	1	0	1	1	1	1	0	0	1	1
2	1	1	0	0	0	0	0	0	1	0	0
3	1	1	1	1	0	1	0	0	1	1	1
4	1	0	0	0	0	0	0	1	0	0	0
5	1	0	0	1	1	1	1	0	0	0	1
6	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	1	0	0	0	0	1	1	0
9	0	1	0	0	0	1	0	0	0	1	1
10	0	0	0	0	0	1	0	0	0	1	0
11	...										
n											

For each proband (Table 8), the record contains several variables:

identification number (ID) of the proband.

a binary response variable (resp) for affection status of the proband: response=1, if the proband is affected with schizophrenia; response=0 if proband is unaffected (*i.e.* a control individual). The proband is not necessarily one of the individuals for whom genotype data (explanatory variables) are available. The patient-to-be-diagnosed is assigned response=0 when added to the reference data set.

a set of explanatory variables: *i.e.* sets of genotypes of mutations found in the schizophrenia patients and family members and controls and family members. The schizophrenia patients and the control individuals are probands (P) as is the patient-to-be-diagnosed. Unaffected family members are the proband's mother (M), father (F), sib(s) (S1, S2, etc.), child(ren) (C1, C2, etc.) or other relatives. Data for affected family members, *e.g.* the proband's mother (MA), father (FA), sibs (SA1, SA2, etc.), children (CA1, CA2, etc.), or other relatives, are entered as separate explanatory variables.

Genetic explanatory variables: Each individual has 0, 1, or 2 copies of any given mutation allele at a given locus. Thus a genotype at each locus contributes two independent explanatory variables. Most of the affected family members will be relatives of schizophrenia probands, but occasionally a relative of an unaffected proband will turn out to be affected with schizophrenia.

Mutations are tabulated as explanatory variables: (see Table 8):

- (i) by the proband or relative in whom they occur, (e.g. P, M, F, S2, C1, MA, FA, SA1, CA1, other);
- (ii) by the specific folate, cobalamin, or pyridoxine gene locus in which they occur (e.g. 1=DHFR locus, 2=MTHFR locus, 3=TCN2 locus, 4=MTR locus, 5=CBS locus, etc.);
- (iii) by the specific mutation within a locus (e.g., 1=the first-designated mutation within a locus, 2=the second-designated mutation within a locus, etc.); and
- (iv) by whether the individual has a single or double dose of the mutation. Thus an explanatory variable P321 records whether the proband has a single dose of the second-designated mutation of the third-designated locus, *i.e.* TCN2. A variable M312 records whether the proband's mother has a double dose of the first-designated TCN2 mutation studied.

In the present hypothetical reference dataset illustrated of genetic explanatory variables (Table 8), partial genotype data for probands, mothers, fathers, sibs and children are given for five gene loci. Not all of the possible explanatory variables are shown. Probands 1-5 are unrelated individuals with the definite clinical diagnosis of schizophrenia; probands 6-10 are unrelated unaffected (control) individuals. Probands 1, 2, 3, 6 and 9 all have a single copy of the first-designated DHFR mutation; proband 3 also has a second copy of that mutation. Probands 1, 3, 5 and 8 all have a single copy of the first-designated mutation at the MTHFR locus; probands 1 and 5 also have a second copy of that mutation. Mothers of probands 1, 3, 5, 9 and 10 all have a single copy of the first-designated DHFR mutation; mothers of probands

1 and 5 also have a second copy of this mutation. Mothers of probands 4 and 7 each have a single copy of the first-designated mutation of TCN2; data for a double dose are not shown. The fathers of probands 2, 3, and 8 each have a single copy of the first designated mutation of CBS; data for a double dose are not shown. The second
 5 (unaffected) sibs of probands 1, 3, 8, 9, and 10 each have a single copy of the first-designated mutation of MTR; data for a double dose are not shown. The first affected children of probands 1, 3, 5, and 9 each have a single copy of the first-designated mutation of DHFR. Other susceptibility loci and mutations can be incorporated in Table 8 in the same fashion *e.g.*, cytokine gene mutations or
 10 polymorphisms, or major histocompatibility complex (MHC) mutations or polymorphisms.

Environmental explanatory variables: If only genetic explanatory variables (genotype data) are used, the maximum predicted probability that the proband is affected with schizophrenia is expected to be approximately about 0.5 in most
 15 populations. When environmental risk factors are included as explanatory variables, the maximum predicted probability that the proband is affected with schizophrenia may approach 1.0. Examples of environmental risk factors for a schizophrenia patient include:

- (1) the proband's dietary folate/cobalamin/pyridoxine intake.
- 20 (2) the proband's circulating levels of folate/cobalamin/pyridoxine.
- (3) the proband's circulating levels of homocysteine, methylmalonic acid, or cystathionine. Elevated levels are indicators of subtle folate/cobalamin deficiency.
- (4) the proband's mother's dietary folate/cobalamin/pyridoxine intake at the time of patient diagnosis, during a pregnancy, or during the pregnancy that produced the
 25 proband.
- (5) the proband's mother's circulating levels of homocysteine, methylmalonic acid, or cystathionine at the time of patient diagnosis, during a pregnancy, or during the pregnancy that produced the proband.

- (6) dietary or circulating folate/cobalamin/pyridoxine or circulating levels of homocysteine, methylmalonic acid, or cystathionine for other family members.
- (7) epidemiological factors related to the proband's gestation and birth, *e.g.* low birth weight or preterm birth, maternal infection, maternal smoking (associated with low plasma folate), season of birth (late winter or spring births are more common in schizophrenia), etc.

Method of Data Analysis

The method exemplified herein is based upon the published guide for the SAS system, but other software can be used. The dataset is analyzed using binary logistic regression to model the response probability, p_i , that the i th proband's affection status is 1, *i.e.* the probability that the i th proband has schizophrenia, given the vector of explanatory variables, x_i . That is:

$$p_i = \text{Prob}(y_i=1|x_i).$$

To do this the logit transformation of p_i is modeled as a linear function of the explanatory variables in the vector, x_i :

$$\text{logit}(p_i) = \log(p_i/[1-p_i]) = \alpha + \beta x_i$$

where: α is the intercept parameter and

β is the vector of slope parameters.

In SAS, the "descending" option is used to model the probability that the response=1, as in the present analysis, rather than response=0.

Outputs of binary logistic regression analysis

After analysis of a dataset, the outputs obtained from SAS include:

- (a) Estimates and standard errors of the parameters (α and β). Using estimates of the intercept parameter (α) and the slope parameter (β) for each environmental or genetic risk factor, the logistic regression equation for the dataset can be written.
- (b) Significance tests of the parameters (*e.g.* Wald chi-square). From the corresponding p -values, the level of significance of each of the environmental or

genetic risk factors is determined. A global significance test of the data with corresponding p-value is also determined.

(c) Odds ratios are given for the slope parameters of each environmental or genetic risk factor. Thus the amount contributed by each environmental or genetic risk factor to the risk of schizophrenia is determined.

(d) The confidence limits for regression parameters and odds ratios are determined.

(e) The predicted probabilities of the observations can be computed, *i.e.* the probability that each individual in the dataset has schizophrenia:

$\alpha \sim$ = estimate of the intercept parameter;

$\beta \sim$ = vector of the estimates of the slope parameters;

x = vector of the explanatory variables;

$p \sim$ = predicted probabilities

$$p \sim = \frac{1}{1 + \exp(\alpha \sim - \beta \sim x)}$$

(f) The model is modified by adding or removing variables until a model is found that best fits the data;

(g) The model is tested for goodness-of-fit. Also, the degree of influence of each specific observation is tested to detect extreme or ill-fitting observations. These may be examples of data entry errors or alternatively, observations that do not fit the present model for schizophrenia.

(h) The probability that a new individual (the patient-to-be-diagnosed) is schizophrenic is then calculated from the final, modified, best fitting regression equation based upon parameters derived from a corrected/modified data set. A simple method of doing this is to add the data for the patient-to-be-diagnosed to the reference data set, a large group of well-studied schizophrenia probands, schizophrenia family members, control probands and control family members for whom data are available for many explanatory variables. A model is created consisting of those informative explanatory variables actually available from the specific patient-to-be-diagnosed and family members participating in the testing. This new combined data set (reference

data set + data from patient-to-be-diagnosed with participating family members) is analyzed by binary logistic regression for the model chosen giving the predicted probability that a proband is affected with schizophrenia for all of the probands including the patient-to-be-diagnosed.

- 5 (i) A classification table is produced from the data set by the "jack knifing" procedure or an approximation to it. This procedure classifies each observation as an event or nonevent based on the model that omits the observation being classified. A classification table sorts observations into percent correct, percent false positives, and percent false negatives at various probability levels and computes
10 sensitivity and specificity.

 (j) The data set used for diagnostic testing is constantly being updated and the regression equation corrected. For example, stratification by geographic residence or geographic origin of ancestors must be considered for some environmental or genetic risk factor.

- 15 For example, in Table 9, entries 34-43 are shown for the data file containing genotypes of 38 schizophrenic probands plus 211 control probands; the first 38 are the affected probands. For individual 302088, the proband is affected ("1"); there is a single dose ("1") of the DHFR mutation but not a double dose ("0") and a single dose ("1") of the MTHFR mutation but not a double dose ("0"). The number 302088
20 identifies the individual whose genotypes are listed; the proband, in this case, is the same individual.

TABLE 9SAS DATAFILE FOR SCHIZOPHRENIA PATIENTS AND CONTROLS

	...
	...
5	34 302086 1 1 0 1 1
	35 302088 1 1 0 1 0
	36 302110 1 1 0 1 0
	37 302111 1 1 0 0 0
	38 302136 1 1 1 1 0
10	39 100001 0 1 0 0 0
	40 100061 0 0 0 0 0
	41 100064 0 1 0 1 0
	42 100067 0 0 0 1 0
	43 100073 0 1 0 0 0
15	...
	...
	...
	...

In Table 10, entries 31-40 are shown for the data file containing genotypes of 35 mothers of schizophrenic probands plus (the same) 211 control probands. For individual 302083, the proband is affected ("1"); there is a single dose of the DHFR mutation ("1") but not a double dose ("0"); there is neither a single ("0") nor a double ("0") dose of the MTHFR mutation. The number 302083 identifies the individual whose genotypes are listed, a mother; the proband, in this case, is a different individual, her affected child.

TABLE 10SAS DATAFILE FOR SCHIZOPHRENIA MOTHERS AND CONTROLS

	...
	...
5	31 302083 1 1 0 0 0
	32 302103 1 0 0 1 0
	33 302104 1 0 0 1 0
	34 302105 1 1 0 1 0
	35 302120 1 0 0 0 0
10	36 100001 0 1 0 0 0
	37 100061 0 0 0 0 0
	38 100064 0 1 0 1 0
	39 100067 0 0 0 1 0
	40 100073 0 1 0 0 0
15	...
	...

In Table 11, entries 11-20 are shown for the data file containing genotypes of 15 fathers of schizophrenic probands plus (the same) 211 control probands. For individual 302084, the proband is affected ("1"); there is a single dose ("1") but not a double dose ("0") of the DHFR mutation; there is both a single ("1") and a double dose ("1") of the MTHFR mutation. The number 302084 identifies the individual whose genotypes are listed, a father; the proband, in this case, is a different individual, his affected child.

TABLE 11SAS DATAFILE FOR SCHIZOPHRENIA FATHERS AND CONTROLS

	...
	...
5	11 302102 1 0 0 0 0
	12 302106 1 1 0 0 0
	13 302115 1 1 0 0 0
	14 302117 1 1 0 0 0
	15 302084 1 1 0 1 1
10	16 100001 0 1 0 0 0
	17 100061 0 0 0 0 0
	18 100064 0 1 0 1 0
	19 100067 0 0 0 1 0
	20 100073 0 1 0 0 0
15	...
	...

In Table 12, entries 9-18 are shown for the data file containing genotypes of 13 unaffected sibs of schizophrenic probands plus (the same) 211 control probands. For individual 302089, the proband is affected ("1"); there is a single dose ("1") but not a double dose ("0") of the DHFR mutation; there is both a single ("1") and a double dose ("1") of the MTHFR mutation. The number 302089 identifies the individual whose genotypes are listed, an unaffected sib; the proband, in this case, is a different individual, the affected sib of individual 302089.

TABLE 12SAS DATAFILE FOR SCHIZOPHRENIA SIBS AND CONTROLS

...						
5	...					
	09	302071		1	1	0 0 0
	10	302073	1	0	0	1 0
	11	302089	1	1	0	1 1
	12	302118	1	1	0	0 0
10	13	302126	1	1	0	0 0
	14	100001	0	1	0	0 0
	15	100061	0	0	0	0 0
	16	100064	0	1	0	1 0
	17	100067	0	0	0	1 0
15	18	100073	0	1	0	0 0
...						
...						

In Tables 9-12 for individual 100061, the proband is unaffected ("0"); there is neither a single dose ("0") nor a double dose ("0") of the DHFR mutation; there is neither a single dose ("0") nor a double dose ("0") of the MTHFR mutation. Since the proband is unaffected, this is a control individual. The number 100061 identifies the individual whose genotypes are listed, as a control individual; the proband, in this case, is the same individual. The identical group of control individuals is used for all four comparisons.

EXAMPLE 2Distribution of Folate Gene Polymorphism Genotypes Among Schizophrenics,
Schizophrenia Parents, Schizophrenia Sibs, and ControlsSummary

- 5 The DNA polymorphism-Diet-Cofactor-Development hypothesis (DDCD hypothesis, described above) postulates that schizophrenia results in part from developmental brain damage sustained *in utero* from the aggregate effect of maternal defects of genes related to important cofactors, *e.g.* folate, cobalamin, pyridoxine, potentiated by a maternal dietary deficiency of these cofactors. The maternal damage to the fetus
- 10 results in part from insufficiency of these cofactors themselves and in part from resulting effects such as immune deficiency and maternal teratogens, *e.g.* hyperhomocysteinemia. Genes from either parent acting in the fetus may modify these damaging effects as outlined in the gene-teratogen model (described above).

- The hypothesis addresses all of the unusual biological and epidemiological features of
- 15 schizophrenia: *e.g.* the decreased amount of grey matter in brain areas, the unusual birth-month effect, the geographical differences in incidence, the socioeconomic predilection, the association with obstetrical abnormalities (low birth weight and prematurity), the decreased incidence of rheumatoid arthritis, and the association with viral epidemics (described above).
- 20 The hypothesis can be supported by finding significant association of sequence variants of folate, cobalamin, or pyridoxine genes with schizophrenia. Folate, cobalamin, and pyridoxine absorption, transport, and metabolism are complex [Rosenblatt, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New York: McGraw-Hill, pp. 3111-3128 (1995); Benton and Rosenberg, In: *The Metabolic and Molecular Bases of Inherited Disease*, Scriver *et al.* (eds), New
- 25 York: McGraw-Hill, pp. 3129-3149 (1995); Whyte *et al.*, *Hypophosphatasia*, In: *The*

- Metabolic and Molecular Bases of Inherited Disease, Scriver *et al.* (eds), New York: McGraw-Hill pp. 4095-4111] with multiple transport proteins, enzymes, and regulatory components. A strong candidate for harboring a mutation predisposing to schizophrenia is the DHFR gene coding for the folate enzyme dihydrofolate reductase. DHFR chemically reduces dietary folate converting it into a form that can enter cellular metabolism. DHFR is also important for DNA synthesis and is known to play a major role in development *in utero*. A novel polymorphic 19 basepair deletion of the DHFR gene has been isolated which could be of functional significance because it affects potential transcription factor binding sites.
- 10 A second candidate is the MTHFR gene, coding for methylenetetrahydrofolate reductase, MTHFR, an important enzyme of folate metabolism. MTHFR was of particular interest because severe deficiency of enzyme activity has been associated with the "schizophrenia" phenotype [Freeman *et al.*, *N. Engl. J. Med.*, **292**:491-496 (1975); Regland *et al.*, *J. Neural Transm. Gen. Sect.*, **98**:143-152 (1994)] and because
- 15 a common mutation, the nt677 C->T transition results in a mutated gene that encodes a heat-labile MTHFR, having decreased enzymatic activity, which in the presence of dietary folate deficiency, causes the plasma homocysteine of homozygotes to become elevated [van der Put *et al.*, *Lancet.*, **346**:1070-1071 (1995); Frosst *et al.*, *Nature Genet.*, **10**:111-113 (1995)]. In adults, hyperhomocysteinemia is known to cause
- 20 vascular disease and to be toxic [Frosst *et al.*, *Nature Genet.*, **10**:111-113 (1995)]. Therefore, homocysteine that crosses the placenta could act as a fetal teratogen during pregnancy. Maternal folate deficiency could also have a more direct teratogenic effect through fetal folate deprivation. These effects could be potentiated by abnormalities of other folate, cobalamin, or pyridoxine genes, even if these
- 25 abnormalities were only minor.

Materials & Methods:

1. *Subjects and Sample Collection:* Patients with schizophrenia and unaffected family members of schizophrenics, were ascertained from patient facilities, patient support

groups, and family support group organizations. Nearly all schizophrenia families had only a single case of schizophrenia. The patients came from different schizophrenia families than the parents and sibs. The controls were unaffected and unrelated individuals not known to be schizophrenic or related to patients with schizophrenia or spina bifida. All subjects were of Caucasian background except two of the schizophrenia patients who were of African American background.

After informed consent was obtained, 20-40 ml of blood was collected into EDTA (purple-top) vacutainers, placed on ice immediately, and transported to the laboratory where plasma, packed red cells, and buffy coat were separated by centrifugation and frozen at -80°C.

2. *Detection of Alleles:* DNA was isolated using the QIAmp column DNA extraction procedure or the QIAGEN Genomic-tip method (QIAGEN, Chatsworth, CA). Alleles for a newly detected polymorphic 19 bp deletion in the dihydrofolate reductase (DHFR) gene were determined by polymerase chain reaction (PCR) amplification of the region surrounding the deletion using specific primers (Fig 1) and direct detection of the PCR products after separation of products on a non-denaturing polyacrylamide gel. A Cetus - Perkin-Elmer 9600 thermocycler was used. Briefly, the PCR reaction contained 200 uM dNTPs, 1.5 mM MgCl₂, 10 pmols of each primer, in 10 ul reaction volume. The PCR conditions used were denaturation at 94°C for 6 min. initially, followed by 35 cycles of 94°C for 55 sec., 60°C for 55 sec., and 72°C for 55 sec. and a final extension at 72°C for 12 min.

Alleles for the 677C->T transition of the methylenetetrahydrofolate reductase (MTHFR) gene were determined by cleavage with the restriction endonuclease, HinfI, of PCR-amplified genomic DNA from blood and separation of the products by non-denaturing polyacrylamide gel electrophoresis [Frosst *et al.*, *Nature Genet.*, 10:111-113 (1995)].

3. *Sequencing the Region Around the DHFR Deletion:* Using the same primers (Figure 1), genomic DNA from individuals with 1,1 and 2,2 genotypes was amplified by PCR and the products sequenced using an ABI PRISM 377 automated sequencer. Restriction sites were identified using the MAP Program in the GCG Package.
- 5 Potential transcription factor binding sites were detected with the TESS program (transcription element search software, URL:<http://agave.humgen.upenn.edu/teess/index.html>).
4. *Data Analysis:* Since the mode of inheritance of schizophrenia is unknown, binary logistic regression was used to test the DHFR deletion allele and the MTHFR
- 10 heat-labile allele as genetic risk factors for schizophrenia. Either the DHFR deletion polymorphism or the MTHFR heat-labile allele could itself be a genetic risk factor for schizophrenia. The genotypes of the two folate gene polymorphisms were used as explanatory variables. Genotypes of schizophrenia patients, parents, or sibs were compared with those of controls.
- 15 Four files were constructed consisting of schizophrenia patients+controls, mothers of schizophrenia patients+controls, fathers of schizophrenia patients+controls, and sibs of schizophrenia patients+controls for input into the SAS System. Each dataset contained 6 variables. In order, these were:
1. six digit identification (ID) number;
 - 20 2. response variable, *i.e.* affection status of the proband
(0=unaffected, *i.e.* control individual; 1=affected, *i.e.* schizophrenia patient);
 3. DHFR mutation-single dose (Ds);
 4. DHFR mutation-double dose (Dd);
 5. MTHFR mutation-single dose (Ms); and
 - 25 6. MTHFR mutation-double dose (Md).
- For mutation data, 0=mutation absent, 1=mutation present.

Results

Alleles of the DHFR 19 bp Deletion Polymorphism: Amplification of the region of intron 1 of DHFR defined by the primers in Figure 1 gave two polymorphic bands of 232 and 213 bp after separation on a non-denaturing polyacrylamide gel (Figure 2).

- 5 Sequencing the PCR products from the two homozygotes showed that they differed by 19 bp (Figure 3). The upper and lower bands (Figure 2), non-deletion allele and deletion allele respectively, were designated alleles 1 and 2 respectively. Comparison with two published sequences showed that allele 1 was identical with one of them [Yang *et al. J. Mol. Biol.* **176**:169-187 (1984)] indicating that allele 2 resulted from a
- 10 19 bp deletion. The other published sequence [Chen *et al. J. Biol. Chem.* **259**:3933-3943 (1984)] was lacking one base pair of allele 1, an A indicated by "*" in Fig 3. It is possible that this shorter reference sequence [Chen *et al. J. Biol. Chem.* **259**:3933-3943 (1984)] resulted from a sequencing artifact.

- Sequences in the 19 bp Deleted Region of DHFR Intron 1:* The 19bp sequence in the
- 15 deleted region (Fig 3) of DHFR intron 1 contained sites for several restriction enzymes including RsaI and ScrFI, and potential binding sites for transcription factors including Sp1, NF-kappaB, CP1 (NF-Y), E2F, ETF and GCF in the 19 base pair region.

- Binary Logistic Regression Analysis:* The number of individuals with each genotype
- 20 of the two polymorphisms among 38 unrelated schizophrenia probands, 35 unrelated mothers of schizophrenia probands, 15 unrelated fathers of schizophrenia probands, 13 unrelated unaffected sibs of schizophrenia probands, and 211 unrelated unaffected control probands is shown in Table 13.

TABLE 13
DISTRIBUTION OF DHFR AND MTHFR MUTATION GENOTYPES
AND ALLELES AMONG CONTROLS, SCHIZOPHRENICS,
AND SCHIZOPHRENIA FAMILY MEMBERS

5		<u>DHFR 19 bp deletion polymorphism:</u>				
	--GenTyp--	-----Schizophrenia-----				---Ctrl---
		P	M	F	S	
	1/1	6 (.16)	10 (.29)	4 (.27)	4 (.31)	56 (.26)
	1/2	22 (.58)	13 (.37)	11 (.73)	8 (.61)	115 (.54)
10	2/2	10 (.26)	12 (.34)	0 (0.0)	1 (.08)	40 (.19)
	total	38 (1.00)	35 (1.00)	15 (1.00)	13 (1.00)	211 (.99)
15		<u>MTHFR 677C->T transition polymorphism:</u>				
	--GenTyp--	-----Schizophrenia-----				---Ctrl---
		P	M	F	S	
	1/1	14 (.37)	16 (.46)	11 (.73)	4 (.31)	103 (.49)
	1/2	18 (.47)	18 (.51)	3 (.20)	8 (.61)	78 (.37)
	2/2	6 (.16)	1 (.03)	1 (.07)	1 (.08)	30 (.14)
	total	38 (1.00)	35 (1.00)	15 (1.00)	13 (1.00)	211 (1.00)

P=schizophrenia patients; M=mothers of schizophrenia patients; F=fathers of
 20 schizophrenia patients; S=unaffected sibs of schizophrenia patients; Ctrl=control
 individuals.

The four data files were analyzed using the logistic procedure of SAS (SAS Institute Inc., 1995) and the "descending" option, which modeled the probability that RESPONSE=1, that is, the probability that the proband was affected with schizophrenia. Note that the proband was not always the individual whose genotype data were used. For example, genotype data for mothers of schizophrenic probands were used to determine the probability that their children, the probands, were affected. Use of the "best" model selection options for logistic analysis in SAS gave the best models for two and three explanatory variables, (Table 14).

Table 14

BINARY LOGISTIC REGRESSION RESULTSGENETIC RISK FACTORMODEL: Ds Dd Ms Md

Odds Ratio (p value)

Schizophrenia Patients

Ds OR(p)	1.937 (.18)
Dd OR(p)	1.263 (.59)
Ms OR(p)	1.775 (.14)
Md OR(p)	0.914 (.86)

Mothers of Schizophrenia Patients

Ds OR(p)	0.630 (.31)
Dd OR(p)	2.653 (.028)*
Ms OR(p)	1.439 (.34)
Md OR(p)	0.143 (.065)

Fathers of Schizophrenia Patients

Ds OR(p)	1.178 (.79)
Dd OR(p)	0.000 (.96)
Ms OR(p)	0.366 (.14)
Md OR(p)	0.841 (.88)

Unaffected Sibs of Schizophrenia Patients

Ds OR(p)	1.104 (.88)
Dd OR(p)	0.337 (.31)
Ms OR(p)	2.688 (.12)
Md OR(p)	0.317 (.29)

Notes For Table 14

DHFR 19 bp deletion: Ds=single dose; Dd=double dose
MTHFR 677C->T mutation: Ms=single dose; Md=double dose

Logistic regression model:

Model with four explanatory variables (Ms, Md, Ds and Dd).

OR(p)=odds ratio and the corresponding p-value for that odds ratio
 determination *=significant at the $p \leq .05$ level.

0.000 odds ratios occurred since none of the fathers of schizophrenia patients had genotype Dd; there was a possibly quasi- complete separation in the sample points; the maximum likelihood estimate may not exist; and therefore validity of the model fit for these odds ratios was questionable.

The comparison of mothers of schizophrenia probands with control probands was statistically significant. Ds was not a significant genetic risk factor. Neither Ms nor Md in mothers was a significant genetic risk factor. However, the p-value for Md decreased and approached significance ($p=.065$) at the $p<.05$ level.

- 5 *Predicted Probabilities of the Various Genotypes:* The "probs predicted" modality of SAS, gave the predicted probability that the proband was affected with schizophrenia (response=1) given genotype data for control probands and schizophrenia patients (probands), mothers of schizophrenia probands, fathers of schizophrenia probands, or sibs of schizophrenia probands. The maximum probabilities obtained are shown in
- 10 Table 15. The highest maximum predicted probability that the proband was affected was obtained for genotype data from mothers of schizophrenia probands, next for schizophrenia probands, next for fathers of schizophrenia probands, and lowest for sibs of schizophrenia probands.

TABLE 15
MAXIMUM PREDICTED PROBABILITY

<u>Model</u>	<u>P</u>	<u>M</u>	<u>F</u>	<u>S</u>
Ds Dd Ms Md 0.24	0.29	0.12	0.11	

Model and explanatory variables are the same as in Table 14.

- Determination of Genotypes Conferring the Highest Risk:* The predicted probabilities that the proband was affected with schizophrenia given specific genotypes of control probands and schizophrenia probands, mothers of schizophrenia probands, fathers of schizophrenia probands, or sibs of schizophrenia probands were determined using the
- 5 model containing all four explanatory variables (Table 16). The predicted probabilities that the proband was affected with schizophrenia were highest for maternal genotypes (Table 15). The maternal genotype with the highest risk was Dd Ms, conferring a probability of 0.29 of schizophrenia in the proband (Table 16). The Dd Ms genotype also gave the highest predicted probability, 0.24, for schizophrenia
- 10 patients.

TABLE 16
PREDICTED PROBABILITIES FOR SPECIFIC GENOTYPES

<u>Model: Ds Dd Ms Md</u>			
<u>Genotype</u>	<u>Predicted</u>	<u>Genotype</u>	<u>Predicted</u>
	<u>Probability</u>		<u>Probability</u>
<u>Schizophrenia Patients:</u>			
Dnull + Mnull	0.07	Ds + Ms	0.20
Dnull + Ms	0.12	Ds + Md	0.19
Dnull + Md	0.11	Dd + Ms	0.24
Ds + Mnull	0.12	Dd + Md	0.23
Dd + Mnull	0.15		
<u>Mothers of Schizophrenia Patients:</u>			
Dnull + Mnull	0.16	Ds + Ms	0.13
Dnull + Ms	0.20	Ds + Md	0.02
Dnull + Md	0.03	Dd + Ms	0.29
Dd + Mnull	0.22	Dd + Md	0.06
Ds + Mnull	0.10		
<u>Fathers of Schizophrenia Patients:</u>			
Dnull + Mnull	0.10	Ds + Ms	0.05
Dnull + Ms	0.04	Ds + Md	0.04
Dnull + Md	0.03	Dd + Ms	0.0
Ds + Mnull	0.12	Dd + Md	0.0
Dd + Mnull	0.0		
<u>Unaffected Sibs of Schizophrenia Patients:</u>			
Dnull + Mnull	0.04	Ds + Ms	0.11
Dnull + Ms	0.10	Ds + Md	0.04
Dnull + Md	0.03	Dd + Ms	0.04
Ds + Mnull	0.04	Dd + Md	0.01
Dd + Mnull	0.02		

Genotypes consist of the same explanatory variables described in Table 14 except that Dnull has no copy of the DHFR deletion and Mnull has no copy of the MTHFR 677C->T variant. Odds ratios of 0.0 were unsatisfactory as described in Table 14.

Discussion

Structure and Function of the DHFR 19 bp Deletion Polymorphism: DHFR polymorphisms have been reported previously [Feder *et al.*, *Nucl. Acids Res.* **15**:5906 (1987); Detera-Wadleigh *et al.*, *Nucl. Acids Res.* **17**:6432 (1989)]. It is known that

- 5 introns are important for message regulation *e.g.*, splicing, or as sites for binding transcription factors. Since the first intron is a relatively common location for regulatory elements, it is possible that the deleted region of DHFR intron 1 could play a role in regulation of DHFR or that the deletion could be a genetic risk factor for schizophrenia because it removes potential transcription factor binding sites.
- 10 Abnormalities of transcription factors and their binding sites may play a role in disease. For example, a polymorphic Sp1 binding site in the collagen type I alpha 1 gene has been associated with reduced bone density and osteoporosis [Grant *et al.*, *Nature Genet.* **14**:203-205 (1996)].

- The Nature of the Putative Folate Genetic Risk Factors for Schizophrenia:* Dd in the
- 15 mother of a schizophrenia proband conferred significantly increased risk of schizophrenia in her child (Table 14). The findings that Dd was a genetic risk factor in mothers but not fathers of schizophrenia probands (Table 15) and that Dd in mothers gave a higher predicted probability than in schizophrenia patients, fathers or sibs (Tables 15 and 16) was consistent with the role of DHFR as a teratogenic locus
 - 20 according to the gene-teratogen model (described above). The finding that a double dose but not a single dose of the DHFR deletion in mothers was a genetic risk factor (Table 16) supported a recessive mode of action in the mother. A teratogenic locus acting in the mother can also act as a modifying or specificity locus in the fetus.

- Neither Ms nor Md in mothers of schizophrenia probands showed statistical
- 25 significance as genetic risk factors for schizophrenia in probands (Table 14). However Md in mothers approached statistical significance ($p=.065$) and appeared to be protective (odds ratio 0.14), while Ms in mothers appeared to increase risk modestly (odds ratio 1.44, $p=.34$).

Role of Genetic and Environmental Factors in Schizophrenia: Since the probability that a schizophrenia co-twin is also affected is reported [Gottesman, *Schizophrenia Genesis*, Schizophrenia Genesis- The Origins of Madness, W.H. Freeman & Co. N.Y.(1991)] to be only 48%, a large part of the risk for schizophrenia would be anticipated to come from environmental factors. Therefore, some controls should have the genetic risk factors for schizophrenia but not be affected with schizophrenia. In the present data set, 6 of 35 schizophrenia mothers and 7 of 38 schizophrenia patients had Dd Ms, the genotype conferring the highest risk, compared with 15 of 211 controls. Since this genotype gave predicted probabilities of schizophrenia in probands of 0.29 and 0.24 respectively, polymorphisms of DHFR and MTHFR could account for a considerable portion of the genetic component of the risk of schizophrenia.

Relation of DHFR to Cytogenetic and Linkage Data for Schizophrenia: As discussed above, the DHFR gene has been located on chromosome 5 at 5q11.2-13.2. A schizophrenia translocation was reported (McGillivray et al.1990; Bassett, 1992) affecting 5q11.2-5q13.3. Also two-point lod scores of 4.64 and 2.29 were found [Sherrington et al., *Nature*, **336**:164-167 (1988)] for the polymorphic markers D5S76 and D5S39 respectively on chromosome 5, in this region [McGillivray et al., *Am. J. Med. Genet.*, **35**:10-13 (1990); Bassett, *Br. J. Psychiatry*, **161**:323-334 (1992)]. Two other linkage studies found small positive lod scores in this region [Coon et al., *Biol. Psychiatry*, **34**:277-289 (1993); Kendler and Diehl, *Schizophr. Bull.*, **19**:261-285 (1993)], but numerous other studies excluded this region under the assumptions and models used [Kendler and Diehl, *Schizophr. Bull.*, **19**:261-285 (1993)]. Recently, new studies have found suggestive evidence for a potential susceptibility locus at a different region of 5q, 5q31 [Schwab et al., *Nat. Genet.* **11**:325-327 (1997)] and 5q22-31 [Straub et al., *Molec Psychiatr.* **2**:148-155 (1997)].

The case-control study presented herein illustrates the usefulness of the DNA polymorphism-Diet-Cofactor-Development and the gene-teratogen models described

above. More importantly, the results presented herein, clearly fail to reject the specific models, *i.e.*, that folate gene polymorphisms can play a role in the etiology of schizophrenia.

5 The present invention is not to be limited in scope by specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

10 Various publications in addition to the immediately foregoing are cited herein, the disclosures of which are incorporated by reference in their entireties.

We Claim:

1 1. A method of generating a genetic reference dataset for use in the
2 determination of the predicted probability for an individual of having a susceptibility
3 for a developmental disorder due to genetic factors or for developing a developmental
4 disorder due to genetic factors or for having offspring that develop a developmental
5 disorder due to genetic factors comprising:

6 (a) collecting a biological sample from a human subject; wherein the
7 human subject is selected from the group consisting of a diagnostic proband, a blood
8 relative of the diagnostic proband, an affected proband, a blood relative of the
9 affected proband, a control proband, and a blood relative of the control proband;
10 wherein the biological sample contains nucleic acids and/or proteins from the human
11 subject;

12 (b) analyzing the nucleic acids and/or proteins from the biological sample;
13 wherein said analyzing results in a partial or full genotype for the alleles of the genes
14 involved in folate, pyridoxine, and/or cobalamin metabolism; wherein said partial or
15 full genotype forms a dataset of genetic explanatory variables for the human subject;
16 and

17 (c) compiling the dataset of genetic explanatory variables from multiple
18 human subjects into a genetic reference dataset.

1 2. A method of generating a genetic and environmental reference dataset for use
2 in the determination of the predicted probability for an individual of having a
3 susceptibility for a developmental disorder due to genetic factors and environmental
4 factors or for developing a developmental disorder due to genetic factors and
5 environmental factors or for having offspring that develop a developmental disorder
6 due to genetic factors and environmental factors comprising:

7 (a) obtaining dietary and epidemiological information for environmental
8 explanatory variables for the human subjects of Claim 1; and

1 (b) combining said environmental explanatory variables with a genetic
2 reference dataset for the human subjects.

1 3. The method of Claim 2 wherein the developmental disorder is selected from
2 the group consisting of schizophrenia, spina bifida cystica, Tourette's syndrome,
3 dyslexia, conduct disorder, attention-deficit hyperactivity disorder, bipolar illness,
4 autism, chronic multiple tic syndrome and obsessive-compulsive disorder.

1 4. A method of generating an environmental reference dataset for use in the
2 determination of the predicted probability for an individual of having a susceptibility
3 for a developmental disorder due to environmental factors or for developing a
4 developmental disorder due to environmental factors or for having offspring that
5 develop a developmental disorder due to environmental factors comprising:

6 (a) obtaining dietary and epidemiological information for environmental
7 explanatory variables for a human subject; wherein the human subject is selected
8 from the group consisting of a diagnostic proband, a blood relative of the diagnostic
9 proband, an affected proband, a blood relative of the affected proband, a control
10 proband, and a blood relative of the control proband; and

11 (b) compiling a dataset of environmental explanatory variables from
12 multiple human subjects into an environmental reference dataset for the human
13 subjects.

1 5. A method of estimating the genetic susceptibility of an individual to have or
2 to develop a developmental disorder comprising:

3 (a) collecting a biological sample from one or more participants; wherein
4 a participant is either the individual or a blood relative of the individual; and wherein
5 the biological sample contains nucleic acids and/or proteins of the participant;

6 (b) analyzing the nucleic acids and/or proteins from the biological sample;
7 wherein said analyzing results in a partial or full genotype for the alleles of the genes

8 involved in folate, pyridoxine, and/or cobalamin metabolism; and wherein said partial
9 or full genotype forms a dataset of genetic explanatory variables for the participants;
10 (c) adding the datasets of genetic explanatory variables obtained from
11 steps (a) and (b) to a genetic reference dataset therein forming a combined genetic
12 dataset;
13 (d) formulating a model comprising the genetic explanatory variables
14 obtained from the participants; and
15 (e) analyzing the combined genetic dataset; wherein a predicted
16 probability for the individual of having or developing a developmental disorder is
17 determined; and wherein the genetic susceptibility of an individual to have or to
18 develop a developmental disorder is estimated.

1 6. The method of Claim 5 wherein said analyzing the combined genetic dataset
2 is performed by binary linear regression.

1 7. The method of Claim 6 further comprising the step of :
2 (f) modifying the model by adding or subtracting a genetic explanatory
3 variable; and re-analyzing the combined genetic dataset by binary logistic regression;
4 wherein a model is chosen that best fits the data.

1 8. The method of Claim 7 further comprising the step of :
2 (g) testing the model for goodness of fit.

1 9. The method of Claim 8 wherein the binary linear regression is performed with
2 the SAS system.

1 10. The method of Claim 5 wherein the developmental disorder is selected from
2 the group consisting of schizophrenia, spina bifida cystica, Tourette's syndrome,
3 dyslexia, conduct disorder, attention-deficit hyperactivity disorder, bipolar illness,
4 autism, chronic multiple tic syndrome and obsessive-compulsive disorder.

1 11. The method of Claim 10 wherein the developmental disorder is schizophrenia
2 and the individual is suspected of being genetically susceptible of having or for
3 developing schizophrenia.

1 12. The method of Claim 11 wherein the individual is suspected of being
2 genetically susceptible for having or for developing schizophrenia because a blood
3 relative has schizophrenia.

1 13. The method of Claim 12 wherein the blood relative is a parent, a sibling, or a
2 grandparent.

1 14. The method of Claim 13 wherein the blood relative is a parent and wherein the
2 parent is the mother of the individual.

1 15. A method of estimating the genetic and environmental susceptibility of an
2 individual to have or to develop a developmental disorder comprising:

3 (a) collecting a biological sample from one or more participants; wherein
4 a participant is either the individual or a blood relative of the individual; and wherein
5 the biological sample contains nucleic acids and/or proteins of the participant;

6 (b) analyzing the nucleic acids and/or proteins from the biological sample;
7 wherein said analyzing results in a partial or full genotype for the alleles of the genes
8 involved in folate, pyridoxine, and/or cobalamin metabolism; and wherein said partial
9 or full genotype forms a dataset of genetic explanatory variables for the participants;

10 (c) obtaining dietary and epidemiological information for environmental
11 explanatory variables for the participants; wherein said information forms a dataset of
12 environmental explanatory variables for the participants;

13 (d) adding the datasets of genetic explanatory variables obtained from
14 steps (a) and (b) and the dataset of environmental explanatory variables of step (c) to
15 a genetic and environmental reference dataset therein forming a combined genetic and
16 environmental dataset;

17 (e) formulating a model comprising the genetic and environmental
18 explanatory variables obtained from the participants; and

19 (f) analyzing the combined genetic and environmental dataset by binary
20 logistic regression;
21 wherein a predicted probability for the individual of having or developing a
22 developmental disorder is determined; and wherein the genetic and environmental
23 susceptibility of an individual to have or to develop a developmental disorder is
24 estimated.

1 16. The method of Claim 15 further comprising the step of :

2 (g) modifying the model by adding or subtracting a genetic or
3 environmental explanatory variable; and re-analyzing the combined genetic and
4 environmental dataset by binary logistic regression; wherein a model is chosen that
5 best fits the data.

1 17. The method of Claim 16 further comprising the step of :

2 (h) testing the model for goodness of fit.

1 18. The method of Claim 17 wherein the binary linear regression is performed
2 with the SAS system.

1 19. A method of estimating the susceptibility of an individual to have offspring
2 that develop a developmental disorder comprising:

3 (a) collecting a biological sample from one or more participants; wherein
4 a participant is either the individual or a blood relative of the individual; and wherein
5 the biological sample contains nucleic acids and/or proteins of the participant;

6 (b) analyzing the nucleic acids and/or proteins from the biological sample;
7 wherein said analyzing results in a partial or full genotype for the alleles of the genes
8 involved in folate, pyridoxine, and/or cobalamin metabolism; and wherein said partial
9 or full genotype forms a dataset of genetic explanatory variables for the participants;

10 (c) adding the datasets of genetic explanatory variables obtained from
11 steps (a) and (b) to a genetic reference dataset therein forming a combined genetic
12 dataset;

13 (d) formulating a model comprising the genetic explanatory variables
14 obtained from the participants; and
15 (e) analyzing the combined genetic dataset by binary logistic regression;
16 wherein a predicted probability for the individual to have offspring that
17 develop a developmental disorder is determined; and wherein the genetic and
18 environmental susceptibility of an individual to have offspring that develop a
19 developmental disorder is estimated.

1 20. The method of Claim 19 further comprising the step of :

2 (f) modifying the model by adding or subtracting a genetic explanatory
3 variable; and re-analyzing the combined genetic dataset by binary logistic regression;
4 wherein a model is chosen that best fits the data.

1 21. The method of Claim 20 further comprising the step of :

2 (g) testing the model for goodness of fit.

1 22. The method of Claim 21 wherein the binary linear regression is performed
2 with the SAS system.

1 23. The method of Claim 22 wherein the individual is a pregnant woman.

1 24. A method of lowering the risk of a pregnant woman who has been determined
2 by the method of Claim 23 to be susceptible to have offspring that develop a
3 developmental disorder comprising administering methylfolate, cobalamin or
4 pyridoxine to the pregnant woman, wherein said administering lowers the risk of the
5 pregnant woman of giving birth to offspring with a developmental disorder.

1 25. A method of determining if any treatment is advisable for a pregnant woman
2 who has been determined by the method of Claim 23 to be susceptible to having
3 offspring that develop a developmental disorder comprising determining the

4 concentration of a risk factor from a tissue sample or body fluid from the pregnant
5 woman; wherein when the concentration of the risk factor is statistically above or
6 below an accepted normal range, treatment is advisable.

1 26. A method of monitoring the effect of the administration of methylfolate,
2 cobalamin or pyridoxine to the pregnant woman of Claim 25, comprising determining
3 the concentration of a risk factor from a tissue sample or body fluid from the pregnant
4 woman; and wherein when the concentration of the risk factor is statistically within
5 an accepted normal range, the treatment is effective.

1 27. The method of Claim 26 wherein the risk factor is selected from the group
2 consisting of homocysteine, folate, and cobalamin.

1 28. The method of Claim 22 wherein the individual is the mate of a pregnant
2 woman.

1 29. A method of treating an asymptomatic individual determined by the method of
2 Claim 23 to be susceptible for developing a developmental disorder comprising
3 administering methylfolate, cobalamin or pyridoxine.

1 30. An isolated nucleic acid encoding a genetic variant of human dihydrofolate
2 reductase comprising a nucleotide sequence having a 19 base-pair deletion spanning
3 nucleotides 540 to 558 of the nucleotide sequence of SEQ ID NO:41.

1 31. The isolated nucleic acid of Claim 30 that has the nucleotide sequence of SEQ
2 ID NO:42.

1 32. An expression vector comprising the nucleic acid of Claim 30 operably
2 associated with an expression control sequence, wherein the nucleic acid is selected
3 from the group consisting of cDNA or genomic DNA.

1 33. A PCR primer that can be used to distinguish SEQ ID NO:42 from the
2 nucleotide sequence selected from the group consisting of SEQ ID NO:41 and SEQ
3 ID NO:45.

1 34. The PCR primer of Claim 33 that comprises 10 to 50 consecutive nucleotides
2 from the nucleotide sequence selected from the group of SEQ ID NO: 41, the
3 complementary strand of SEQ ID NO: 41, SEQ ID NO:42, the complementary strand
4 of SEQ ID NO: 42, SEQ ID NO:45, and the complementary strand of SEQ ID NO:
5 45.

1 35. The PCR primer of Claim 34 wherein the 10 to 50 consecutive nucleotides are
2 from nucleotides 350 to 530 of SEQ ID NO:41.

1 36. The PCR primer of Claim 35 having the nucleotide sequence of 5'-CTA AAC
2 TGC ATC GTC GCT GTG-3' (SEQ ID NO:38).

1 37. The PCR primer of Claim 36 wherein the 10 to 50 consecutive nucleotides are
2 from the complementary strand of nucleotides 550 to 850 of SEQ ID NO:41.

1 38. The PCR primer of Claim 37 having the nucleotide sequence of 5'-AAA AGG
2 GGA ATC CAG TCG G-3' (SEQ ID NO:39).

1 39. An isolated nucleic acid that hybridizes under standard hybridization
2 conditions to a nucleic acid having the nucleotide sequence
3 ACCTGGGCGGGACGCGCCA (SEQ ID NO:40) or a sequence complementary to
4 SEQ ID NO:40; wherein said isolated nucleic acid consists of 12 to 48 nucleotides.

1 40. An isolated nucleic acid that hybridizes to the nucleotide sequence of SEQ ID
2 NO:42, but not to the nucleotide sequence of SEQ ID NO:41; when said hybridizing
3 is performed under identical conditions.

- 1 47. The method of Claim 46 wherein said determining is performed by a method
- 2 selected from the group consisting of PCR, special PCR, RT PCR, RFLP analysis,
- 3 SSCP, and FISH.

09577266-032300

Abstract

The present invention discloses a novel method for identifying an individual who may be susceptible to develop a developmental disorder. In one particular example, an individual is identified who is genetically susceptible to becoming schizophrenic. In
5 addition, the present invention discloses a novel method for identifying individuals who are genetically susceptible to have offspring with a developmental disorder. Methods of diagnosing, preventing and treating developmental disorders such as schizophrenia are also provided.

00577206-052370

Primers for PCR Amplification the DHFR Deletion Polymorphism Region

Forward primer: 5'-CTA AAC TGC ATC GTC GCT GTG-3'

Reverse primer: 5'-AAA AGG GGA ATC CAG TCG G-3'

601-1-057 N (Sheet 2 of 5)

Genotypes of the DHFR 19 bp Deletion
by Non-denaturing Polyacrylamide Gel Electrophoresis

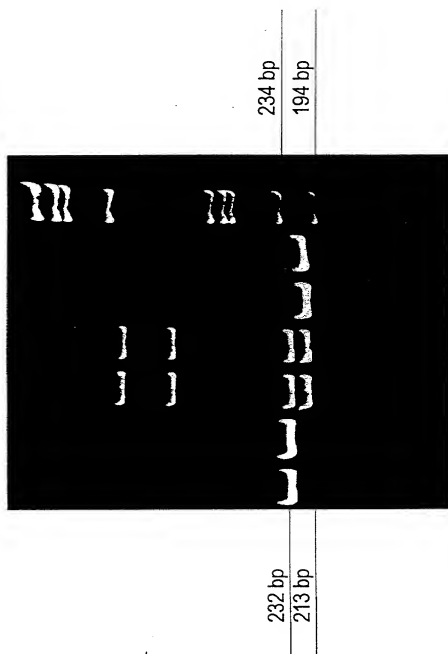


Figure 2

Sequences of PCR Amplification Products
in the Region of the DHFR Deletion Polymorphism Region

Allele 1	GCTGCCACGGTGGGGTAACCTGGGGGGACGGCCAGGCGGACTCCCGGCGAGA
Allele 2	GCTGCCACGGTGGGGT.....GGCGGACTCCCGGCGAGA

Figure 3

601-1-057 N (Sheet 4 of 5)

1 CTGCAGCGCC AGGGTCCACC TGGTCGGCTG CACCTGTGGA GGAGGAGGTG
 51 GATTTCAGGC TTCCCGTAGA CTGGAAGAAT CGGCTCAAAA CCGCTTGCTT
 101 CGCAGGGGCT GAGCTGGAGG CAGCGAGGCC GCCCGACGCA GGCTTCCGGC
 151 GAGACATGGC AGGGCAAGGA TGGCAGCCCG GCGGCAGGGC CCGGCAGGGA
 201 GCGCGAACCC GCGGCCCGCAG TTCCAGGCGC TCTCGGGGCG CGAGCACGCC
 251 GCGACCTTGC GTGCGCCGGG GCGGGGGGGC GGGGCCTCGC CTGCACAAAT
 301 AGGGACGAGG GGGCGGGGCG GCCACAATTT CGCGCCAAAC TTGACCCGCG
 351 GTTCTGCTGT AACGAGCGGG CTCGGAGGTC CTCCTCGCTGC TGT**CATGG**TT
 401 GGTTTCGTAA ACTGCATCGT CGCTGTGTCC CAGAACATGG GCATCGGCAA
 451 GAACGGGGAC CTGCCCTGGC CACCGCTCAG GTATCTGCCG GGCCGGGGCG
 501 ATGGGACCCA AACGGGCGCA GGCTGCCCCC GGTCGGGGTA CCTGGGCGGG
 551 ACGCGCCAGG CCGACTCCCG GCGAGAGGAT GGGGCCAGAC TTGCGGTCTG
 601 CGCTGGCAGG AAGGGTGGGC CCGACTGGAT TCCCTTTTC TGCTGCGCGG
 651 GAGGCCAGT TGCTGATTTC TGCCCGGATT CTGCTGCCCC GTGAGGTCTT
 701 TGCCCTGCGG CGCCCTCGCC CAGGGCAAAG TCCAGCCCTT GGAGAAAACA
 751 CCTCACCCCT ACCCACAGCG CTCCTTTTGT CAGGTGCCCT AGAGCTCGAG
 801 CCCAAGGGAT AATGTTTCGA GTAACGCTGT TTCTCTAACT TGTAGGAATG
 851 AATTCAGATA TTTCCAGAGA ATGACCACAA CCTCTTCAGT AGAAGGTAAT
 901 GTGGGATTAA GTAGGGTCTT GCTTGATGAA GTTTACCAGT GCAAAATGTA
 951 GTTAAATGGA AAGTTTTCG TGTTAATCTG GGACCTTTTC TCTTATTATG
 1001 GATCTGTATG ATCTGTATGC AGTTCCCAAG GTTCATTTAC CATTATTAAA
 1051 AAATTTTTGT CTTAGAAATT TTATGTATGT CAACGCAGCA GCAAAATTATC
 1101 AGGCATGGGG CAGAATTGGC AACTGGGTGG AGGCTTCGGT GGAGGTTAGC
 1151 ACTCCGAAG GAAAACAGAG TAGGCCTTTG GAACAGCTGC TGGGAAGAGAT
 1201 AAGGCCTGAA CAAGGGCAGT GGAGAAGAGA GGGTAAAAAT TTTTAAAGGT
 1251 TACATGACCC TGGATTTTGG AGATC

"

Figure 4A

601-1-057 N (Sheet 5 of 5)

1 CTGCAGCGCC AGGGTCCACC TGGTCGGCTG CACCTGTGGA GGAGGAGGTG
 51 GATTTTCAGGC TTCCCGTAGA CTGGAAGAAT CGGCTCAAAA CCGCTTGCCT
 101 CGCAGGGGCT GAGCTGGAGG CAGCGAGGCC GCGCGACGCA GGCTTCCGGC
 151 GAGACATGGC AGGGCAAGGA TGGCAGCCCG GCGGCAGGGC CCGGCAGGGA
 201 GCGCGAACCC GCGGCCGCAG TTCCAGGCG TCTGCGGGCG CGAGCACGCC
 251 GCGACCCTGC GTGCGCCGGG GCGGGGGGGC GGGGCCCTGC CTGCACAAAT
 301 AGGGACGAGG GGGCGGGGCG GCCACAATTT CGCGCCAAAC TTGACCGCGC
 351 GTTCTGCTGT AACGAGCGGG CTCGGAGGTC CTCCCGCTGC TGTCATGGTT
 401 GGTTCGCTAA ACTGCATCGT CGCTGTGTCC CAGAACATGG GCATCGGCAA
 451 GAACGGGGAC CTGCCCTGGC CACCGCTCAG GTATCTGCCG GGCCGGGGCG
 501 ATGGGACCCA AACGGGCGCA GGCTGCCAC GGTCGGGGT
 551 GG CCGACTCCCG GCGAGAGGAT GGGGCCAGAC TTGCGGCTCTG
 601 CGCTGGCAGG AAGGGTGGGC CCGACTGGAT TCCCTTTTC TGCTGCGCGG
 651 GAGGCCAGT TGTGATTTC TGCCCGGATT CTGCTGCCCG GTGAGGTCTT
 701 TGCCCTGCGG CGCCCTCGCC CAGGGCAAAG TCCAGCCCT GGAGAAAACA
 751 CCTCACCCCT ACCCACAGCG CTCGTTTGT CAGGTGCCTT AGAGCTCGAG
 801 CCCAAGGGAT AATGTTTCGA GTAACGCTGT TTCTCTAACT TGTAGGAATG
 851 AATTACAGATA TTCCAGAGA ATGACCACAA CCTCTTCAGT AGAAGGTAAT
 901 GTGGGATTAA GTAGGGTCTT GCTTGATGAA GTTACCAGT GCAAATGTTA
 951 GTTAAATGGA AAGTTTTCG TGTTAATCTG GGACCTTTTC TCTTATTATG
 1001 GATCTGTATG ATCTGTATGC AGTTCCCAAG GTTCATTAC CATTATTAAA
 1051 AAATTTTGT CTTAGAAATT TTATGTATGT CAACGCACGA GCAAATTATC
 1101 AGGCATGGGG CAGAATTGGC AACTGGGTGG AGGCTTCGGT GGAGGTTAGC
 1151 ACTCCGAAAG GAAAACAGAG TAGGCCTTTG GAACAGCTGC TGAAGAGAT
 1201 AAGGCCTGAA CAAGGGCAGT GGAGAAGAGA GGGTAAAAAT TTTTAAAGST
 1251 TACATGACCC TGGATTTTGG AGATC

Figure 4B

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As below named inventors, we hereby declare that:

Our residence, post office address and citizenship are as stated below under our names.

We believe that we are the original, first and joint inventors of the subject matter which is claimed and for which a patent is sought on the invention entitled

**METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
DEVELOPMENTAL DISORDERS DUE TO A COMBINATION OF
GENETIC AND ENVIRONMENTAL FACTORS**

the Specification of which

[X] is attached hereto
[] was filed on _____
as Application Serial No. _____
and was amended on _____ (if applicable).

We hereby state that we have reviewed and understand the contents of the above-identified Specification, including the claims, as amended by any amendment referred to above.

We acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, 1.56(a).

We hereby claim foreign priority benefits under Title 35, United States Code, §119 of any provisional application filed in the United States in accordance with 35 U.S.C. §1.119(e), or any application for patent that has been converted to a Provisional Application within one (1) year of its filing date, or any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

PRIOR FILED APPLICATION(S)

<u>APPLICATION NUMBER</u>	<u>COUNTRY</u>	<u>(DAY/MONTH/YEAR FILED)</u>	<u>PRIORITY CLAIMED</u>
60/136,198	USA	25 MAY 2000	YES

We hereby claim the benefit under Title 35, United States Code, §120 of any United States application listed below, and, insofar as the subject matter of each of the claims of this application is not disclosed in any prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a), which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION NO. _____	FILING DATE (DAY/MONTH/YEAR)	STATUS - PATENTED, PENDING, ABANDONED _____
--------------------------	---------------------------------	--

We hereby appoint as our attorneys or agents the following persons: Jack Matalon, (Attorney, Registration No. 22,441); Stefan J. Klauber (Attorney, Registration No. 22,604); David A. Jackson (Attorney, Registration No. 26,742); Donald J. Cox, Jr. (Attorney, Registration No. 37,804); Michael D. Davis (Attorney, Registration No. 39,161); William C. Coppola (Attorney, Registration No. 41,686); Mark S. Cohen (Attorney, Registration No. 42,425); Steven B. Stein (Attorney, Registration No. 43,159); Christine E. Dietzel (Agent, Registration No. 37,309); Yuan Kong (Agent, Registration No. P43,728) and Michael A. Yamin (Agent, Registration No. P44,414), said attorneys or agents with full power of substitution and revocation to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Please address all correspondence regarding this application to:

DAVID A. JACKSON, ESQ.
KLAUBER & JACKSON
411 HACKENSACK AVENUE
HACKENSACK, NEW JERSEY 07601

Direct all telephone calls to David A. Jackson at (201) 487-5800.

We hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF FIRST JOINT INVENTOR: William G. Johnson, MD

COUNTRY OF CITIZENSHIP: United States

FULL RESIDENCE ADDRESS: 91 Stewart Road
Short Hills, NJ 07078

FULL POST OFFICE ADDRESS: 91 Stewart Road
Short Hills, NJ 07078

SIGNATURE OF INVENTOR _____

DATE _____

FULL NAME OF SECOND JOINT INVENTOR: Edward Scott Stenroos

COUNTRY OF CITIZENSHIP: United States

FULL RESIDENCE ADDRESS: 317 Ann Street, 2nd Floor
Harrison, NJ 07029

FULL POST OFFICE ADDRESS: 317 Ann Street, 2nd Floor
Harrison, NJ 07029

SIGNATURE OF INVENTOR _____

DATE _____

05/17/2016 15:25:00

SEQUENCE LISTING

<110> Johnson, William G.
Stenroos, Edward S.

<120> METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
DEVELOPMENTAL DISORDERS

<130> 601-1-057N

<140> UNASSIGNED

<141> 2000-05-23

<150> 60/136,198

<151> 1999-05-25

<160> 46

<170> PatentIn Ver. 2.0

<210> 1

<211> 2187

<212> DNA

<213> Homo sapiens

<400> 1

```

gccatggtga acgaagccag aggaacacgc agcctcaacc cctgcttgga gggcagtgcc 60
agcatgtgca gtgagagctc caaagatagt tgcagatgtt ccaccccggg cctggaccct 120
gagcgccatg agagactccg ggagaagatg aggcggcgat tggaaatctgg tgacaagtgg 180
ttctccctgg aattcttccc tctcgaact gctgagggag ctgtcaatct catctcaagg 240
tttgaccgga tggcagcagg tggccccctc tacatagacg tgaccttgga cccagcaggt 300
gaccttggtc cagacaagga gacctcctcc atgatgatcg ccagcaccgc cgtgaactac 360
tgtggcctgg agaccatcct gcacatgacc tgctgcctgc agcgctcggg ggagatcacg 420
ggccatctgc acaaagctaa gcagctgggc ctgaagaaca tcatggcgtc gcggggagac 480
ccaataggtg accagtgagg agaggaggag ggaggcttca actacgcagt ggacctggtg 540
aagcacatcc gaagtgtggt ttgtgactac tttgacatct gtgtggcagg ttaccceaaa 600
ggccaccccg aagcaggggag ctttgaggct gacctgaagc acttgaagga gaagtggtct 660
gcgggagccg attttcatcat caccgagctt ttctttgagg ctgacacatt cttccgcttt 720
gtgaaggcat gcaccgacat gggcatcact tgcacctatg tccccggat ctttcccatc 780
cagggtcacc actcctctcg gcagcttgtg aagctgtcca agctggaggt gccacaggag 840
atcaaggagc tgatgtgacc aatcaaagac aacgatgctg ccatccgcaa cttgggcatc 900
gagctggccg tgagcctgtg ccaggagctt ctggccagtg gcttggtgcc aggcctccac 960
ttctacacc tcacccgcga gatggctacc acagaggtgc tgaagcgctt ggggatgtgg 1020
actgaggacc ccagcgctcc cctaccctgg gctctcagtg cccaccccaa gcgcgagag 1080
gaagatgtac gtcccattct ctgggctctc agaccaaaaga gttacatcta ccgtaccacg 1140
gagtgaggac agttccctaa cggccgctgg ggcaattctc cttccctcgc ctttggggag 1200
ctgaaggact actacctctt ctacctgaag agcaagtccc caaggagga gctgctgaag 1260
atgtgggggg aggaagctgac cagtgaagca agtgctttg aagtctttgt tctttacctc 1320
tcggggagac caaacccgaa tggtcacaaa gtgacttgcc tgccctggaa cgatgagccc 1380
ctggcgctcg agaccagcct gctgaaggag gagctgctgc gggatgaacc ccaggggcatc 1440
ctcaccatca actcacagcc caacatcaac ggggaagcgt cctccgaccc catcgtgggc 1500
tgggggccca gcgggggcta tgtcttccag aaggcctact tagagttttt cacttccgcg 1560
gagacagcgg aagcacttct gcaagtgtg aagaagtacg agctccgggt taattaccac 1620
ctgttcaatg tgaagggtga aaacatcacc aatgccctg aactgcagcc gaatgctgtc 1680
acttggggca tcttccctgg gcgagagat atccagccca ccgtagtggg tcccgtcagc 1740
ttcatgttct ggaaggacga ggcctttgac ctgtggattg agcgatggg aaagctgtat 1800
gaggaggagt ccccgctccc caccatcac cagtacatcc acgacaacta cttctctgtc 1860

```

SEQUENCE LISTING

aacctggttg	acaatgactt	ccactggac	aactgcctct	ggcaggtggt	ggaagacaca	1920
ttggagcttc	tcaacaggcc	caccagagat	gcgagagaaa	cggaggctcc	atgaccctgc	1980
ctctcgacgc	ctctcgcttg	agccactctc	ctccgcctct	ctctctccac	agctctgctt	2040
ctcttgggaa	ctccactctc	cttctgtgtc	ctccaccctc	ggcctccact	ccccaccctg	2100
acaatggcag	ctagacttga	gtgaggcttc	caggctcttc	ctggacctga	gtcggcccca	2160
catgggaacc	tagtactctc	tgctctca				2187

<210> 2
 <211> 7122
 <212> DNA
 <213> Homo sapiens

<400> 2						
gcgcgtgtct	ggctgctag	ccgacaccaa	ggaactggcg	ggtaccgcgg	aagaaagcac	60
gtgtctccagc	agttgccgcg	cccagccccc	agagaggccc	tagggcgctg	cgggctttcg	120
gggtccgcagc	tccccccgcg	acgcgagcca	acgggaggcg	tcaaaagacc	cgggccttgt	180
gtggcaggct	cgctctggcg	tggctggcgt	ggcccttggc	cgctgcacc	tgtggagagc	240
acgtctcttc	tgccgcgcgc	tctgcgcgag	gaggagactc	gacaacatgt	caccgcgcgt	300
ccaagacctg	tcgcaacccc	aaggtctgaa	gaaaaccctg	cgggatgaga	tcaatgccat	360
tctgcagaa	aggattatgg	tgctggatgg	agggatgggg	accatgatcc	agcgggagaa	420
gctaaacgaa	gaacactctc	gaggtcagga	atttaagat	catgccagcc	cgctgaaaag	480
caacaatgac	attttaagta	taactcagcc	tgatgtcatt	taccaaattc	ataaggaata	540
cttgcctgct	ggggcgagata	tcattgaaac	aaatactttt	agcagcacta	ctatcgccca	600
agctgactat	ggccttgaac	acttggccta	ccgagtgaac	atgtgctctg	caggagtggt	660
cagaaaaact	gccgagagg	taactctcca	gacaggaatt	aagaggtttg	tggcagggcg	720
tctgggtccg	actaataaga	cactctctgt	gtcccatctc	gtgaaaaggg	cggattatag	780
gaacatcaca	tttgatgagc	ttgttgaagc	ataccaagag	caggccaaag	gactcttgga	840
tggcgggttg	gatactctac	tcattgaaac	tatttttgat	actgccaatg	ccaaggcagc	900
cttgttttga	ctccaaaatc	tttttgagga	gaaatatgct	ccccggccta	tctttatttc	960
agggacgac	gttgataaaa	gtggggcgga	tctttccgga	cagacaggag	agggatttgt	1020
catcagcgtg	tctcatggag	aaccactcta	cattggatta	aattgtgctt	tgggtgcagc	1080
tgaaatgaga	cccttttatg	aaataattgg	aaaatgtaca	acagcctatg	tcctctgtta	1140
tcccaatgca	ggctctccca	acacttttgg	tgactatgat	gaaacgcctt	ctatgatggc	1200
caagcactca	aaggattttg	ctatggatgg	cttggctaat	atagtgtgag	gatgtgtgtg	1260
gtcaacacca	gatcatatca	gggaaattgg	tgaagctgtg	aaaaatgtga	agcctagagt	1320
tccacctggc	actgcttttg	aaggacatac	gttactgtct	ggtctagagc	cttccagagt	1380
tgagaccgtac	accaactttg	ttaacattgg	agagcgtctg	aatgttgcag	gatcaaggaa	1440
gtttgtctaaa	ctcatcatgg	caggaaaacta	tgaagaagcc	ttgtgtgttg	ccaaagtgtca	1500
ggttgaaatg	ggagcccagg	tggtggatgt	caacatggat	gatggccatg	tagatgttcc	1560
aagtgcgaat	accagatttt	gcaacttaat	tgcttccgag	ccagacatcg	caaagggtacc	1620
tttgtgcatac	gactctccca	attttgcgtg	gattgaaagt	gggttaaaag	gctgcctcaag	1680
gaagtgcatt	gtcaatagca	ttagtctgaa	gggaaggagag	gcagactctc	tggagaaggc	1740
cagggaagat	aaaaagtatg	gagctgctat	ggtggctcatg	gactttgatg	aagaaggaca	1800
ggcaacagaa	acagacacaa	aaatcagagt	gtgcaccccg	gcctaccatc	tgcttgtgaa	1860
aaaaactgggc	tttaatccaa	atgacattat	ttttgacctc	aatatcttaa	catctgggag	1920
tggaatggag	gaacacaaat	tgatgcctat	taatttttatc	catgcacaaa	aagtcattaa	1980
agaaacattta	ccctggagcca	gaataaagtg	aggtctcttc	acttctgctc	tctctcttcc	2040
aggaatggaa	gccatctcag	aagcaatgca	tggggttttc	ctttaccatg	caatcaagtc	2100
tggcatggac	atggggatag	tgaatgtctg	aaacctccct	gtgtatgatg	atatccataa	2160
ggaactcttc	cagctctctg	aagatctcat	ctggaataaa	gacctcgagg	cacttgagaa	2220
gctctttacgt	tatgccccaga	ctcaaggcac	aggagggaag	aaagtcatct	agactgatga	2280
gtggagaata	ggccctctgc	aagaacgcct	tgagtatgcc	cttgtgaagg	gcatgaaaga	2340
acatattatt	gaggatactc	aggaagccag	gttaaaccaa	aaaaatatc	cccgacctct	2400
caataataat	gaaggaccct	tgatgaatgg	aatgaaaatt	gtgtgtgac	tttttggagc	2460
tggaaaaaat	tttctcacct	aggttataaa	gtcagcccg	gttatgaaga	aggtctgttg	2520
ccaccttatc	ctcttcctag	aaaaagaagc	agaagaagcc	agagctctta	acggcaacgt	2580
agaaagaatg	gacctcttacc	agggcaccat	cgtgtggccc	actgtttaag	gcgacgtgca	2640

gacataggc	agaagacag	ttaggtagt	cctggctgc	aataatttc	gagttattg	2700
ttdaggatg	atgactccat	gtgataagat	actgaaagct	gctcttgacc	acaaagcaga	2760
tataattggc	ctgtcaggac	ctcatcactc	ttccctggat	gagctgattt	ttgtttgccaa	2820
ggaaattcag	agattagtta	taaggaattc	atctgtgat	ggagaggacaa	ccactctcaa	2880
aacccacaga	gcagttaaaa	tagctccgag	atcacagtca	ctgttaactc	attgtctgga	2940
cgcgtccaaq	agtggtggtg	tgtgttccca	gctgttagat	gaaaattctaa	agatgatata	3000
ctttgaggaa	atcatggaaq	aatatgaaga	tattagacag	gaccattatg	agttctccaa	3060
ggagaggaga	tacttaccct	taagtcaagc	cagaaaaagt	ggtttccaaa	tggattggct	3120
gtctgaacct	caccagatga	agcccacgtt	tattgggacc	caggttcttt	aagactatga	3180
ccctgcagaq	ctgtgtgaat	acattgtagc	gaagcctttc	tattgtgctt	ggcagctccg	3240
gggcaatcag	ccgaattcag	gctttcccaa	gatatttaac	gacaaaacag	taggtggaga	3300
ggccaggaag	gtctacgatg	atgccccaaa	tatgtcgaac	actcagatba	gtcaaaaagc	3360
actcccgccg	cgggctgtgt	tgtgggtctg	ccgcagcaag	agatctcaag	acgacatcta	3420
ctcgtacgca	gaggtctcgt	tgcgccagtg	tgcagagccc	atgtcaacct	tctatggggt	3480
aaggcaacag	gtcgataagg	actctcgagc	cagggaacca	tactactctc	tctcagctat	3540
ctatcgctcc	tgtcgaatg	gcacatcgta	ctactctggc	ctgtttgcgc	tttgcgtgct	3600
tggggtagaa	gagctgagca	aggcctatga	ggatgatggt	gacgactaca	gcagcatcat	3660
ggctcaaggc	ctggggagag	ggctgtcgga	ggcctttgca	gaagagctcc	atgaaagagt	3720
tccgcgacga	ctgtgggctc	ctgtgtcgag	tgcagctgag	ctgcgtcgag	actcgcgcag	3780
gtctcgctac	aaggccatcc	ctgcggcgctc	tggctacccc	acgacgcccg	acccaacaga	3840
gaagctccac	atgtggagac	tccgaagcat	cgcagcagct	acaggcatcta	ggttaacaga	3900
atcatatgca	ctgcgacctg	cttcagcagct	ctcagccgct	tacttctctc	atttgaagtc	3960
caaatatttt	gtgtgtggga	agatttccaa	ggatcaggtt	gaggtattat	catttaggaa	4020
gaacatatct	gtgcttgagg	ttgagaatgt	gttgtgacct	attttgggat	atgacacaga	4080
ctaaactttt	ttttttttgc	cttttttttt	ctttgatctc	cttaaggaaa	tacaacctag	4140
ggtgccttaa	aataaacac	aacaaaaaac	ctgtgtgcat	ctggtcgaca	cttacctgct	4200
cttgggtttc	gaagactatt	tagtggaaac	ttgttagaga	gcaggtgtgt	ccgtcagctg	4260
tctgtaaaaa	ggcctgtttt	tttttgggac	ttcgtggaag	agcagctgagc	aggtgtctgt	4320
tggtttccct	gtgtccctct	agatggggac	agactcaga	gagagtgcgt	ttgatttcaa	4380
agcaagtcga	ctcgtctttt	cttgtttttt	cagtggaaat	taggaggctt	ctatgctgtc	4440
tttttttctc	ctttagaagaa	aagctcgaat	ctgagttaag	tagagaatgt	tgacctctgt	4500
acaaaatagt	actgtgaaaa	atggggcatt	taaatctgaa	tggttataac	agttgattat	4560
gcgggggaag	gtgtagctct	gttctcttgc	gaagcagctc	ttttctaaag	gtcggaacta	4620
atggctcgag	aactcctttt	ggcaaaaggc	atgcgctcac	ttctgtgctt	tcagaaaacac	4680
tgaagccatt	tgcgccagtg	tggtcaagca	gccatgcttt	ctgggctgtt	tcgtctctcc	4740
ataattttat	atttccctag	ccctgaggaa	acaaaagaa	aagttagagg	gaaggttact	4800
gttaaggggtg	gtttaaactt	tttttgtttt	gttttgtttt	ggtttttttt	tttttagaca	4860
gagctctgct	ctgtcgccca	ggctggagtg	caggggcgca	actctggctc	atagcaagct	4920
ccgcctctct	ggttcatcgt	atctctctct	ctcagcctcc	agatgagctg	ggactacagc	4980
tgcaccaac	cacaccggcg	ttaatttttt	tgtttttaca	aaatacgaat	aagtagagac	5040
aggtattctc	tgtgttagcc	aggatggtct	tgatctcccg	acctcgtgat	ctgcacctac	5100
cagcctccca	aaatctctgg	attacagctt	tgagcggcgt	agcctggccg	gttaacatct	5160
tttaattgtt	tcaggagttt	acagaggtct	cagctgggct	ctgtatatcc	gtcgggagtt	5220
ggacaagtgt	gcagcataaa	gtcactcaat	tcttaccatt	ttatccctct	caattctcaa	5280
tatatctagt	aatgaagaat	gggtccacca	ctcaagcaac	aagctctaaa	ctcaacctgt	5340
tactcttttt	ctttgatgat	tcaggttat	tcaaaaattt	gaactcaaaa	tatacatcaa	5400
tctactattc	agatggtgtg	ggcaatagtg	aggagaagat	aacctgtgag	tctcgtgttg	5460
attcgaagga	tgaagcgggt	gaagcaatgg	aggaaacaat	gaagaacctat	ctttgttctc	5520
gaataggaat	attcaagatt	atcaagggat	caggtctctc	aaaattgatc	tatgatttta	5580
ataccatttt	caatgggaaat	tccaacagat	ttttatgaat	gaacaacaga	gggttttata	5640
tggagtacga	aaggactctaa	aattaccaaa	tgctttctaa	tgtgaaggag	aggtttggga	5700
caccctcctc	atggtataacc	aagttttttt	gtcaagacag	tgctatctgt	cagaggtagg	5760
cattctaggc	aggggaacaa	ataaagggct	tagaaaactt	cccgctcata	tgttgacctc	5820
tgcaaaataa	ctcgttgaca	tggcaagtca	gtggggacag	gaaggaccac	tccctaahta	5880
atcccaagca	aatgctcatt	catgtgggaa	aaaaagaaat	ttactttctc	ctcaacctac	5940
ctgtgtataa	tgttccaaaat	tgttaagggc	tttaatacaa	agaacaaaata	tgtcagtgtg	6000
tttgatgaaa	aaagccttag	ggcaggaagaa	aaatctctga	gacataaaagt	agatcaatca	6060

aaggacaaga	tgggttaagtc	aattctgtta	aaactcaagg	cttatattaa	gcaaacactt	6120
gaagtggaga	gatgatccac	aacttgagaa	gacatttata	atacaataa	ctgatgaagg	6180
atcataatc	acaaatatag	agaattccta	tttaaaaaaa	tagaaaaata	gtgaagacta	6240
cacaagagga	aatagggcct	ttaataaaat	agatgttctg	tagcatgtgt	cagggaataa	6300
tgaacttaga	ccacaatgag	atccattttt	atatccataa	gatttgcata	ggttgggtct	6360
gacagtacca	gttgttagat	ctgtagggac	ttgtacaaca	ttgtggatgt	gtaaacaggc	6420
accactgctt	taaaaaacaa	ttatccctta	cagacttgaa	catttgcaga	cgttatgac	6480
ttgtctccaa	ctcccactgt	tatgtccagc	aaactcttgc	atgtggccac	tagggagaa	6540
gtgtaagaat	gttcatagtt	acatatattt	aatagttaat	aactggaaaa	agtgaatagt	6600
atgtctgtct	acaggaataa	aggtgaataa	ttagatatat	atatctatct	tacgggatat	6660
tattcagtag	tggaaaatgag	tgaactacag	ctataacctca	caataagaat	gaatctcaga	6720
aaatatlaag	gaaaaaagca	agtttgaa	gaccacatgg	ggcgtactact	ttttattggg	6780
cccaaaaaaca	agcaaaacca	aagaatatgt	agtcctaagca	tacgtataca	ataaaactat	6840
gctattaaaa	aaaaaaggta	actgataaac	caaaattgag	catagtaatt	accacagaaa	6900
ggaggaagtg	gaagggacag	gagcacatag	gtagatgccca	agttatgcag	ctgttcttgt	6960
tcctccttgt	aggcttcaaa	gtgtttacta	tatgttatta	atacatatta	ctttataact	7020
aatagataaac	agttttttac	atatttaata	tgttctactt	aaatatatta	taaaaaataa	7080
aggcaaaagt	gaatgtttaa	aaaaaaaaaa	aaaaaaaaaa	aa		7122

<210> 3
 <211> 564
 <212> DNA
 <213> Homo sapiens

<400> 3						
atgggttggtt	cgctaaactg	catcgtcgct	gtgtcccaga	acatgggcat	cggaagaac	60
ggggacctgc	cctggccacc	gctcaggaat	gaattcagat	attccagag	aatgaccaca	120
acctcttcag	tagaaggtaa	acagaalctg	gtgattatgg	gtaagaagat	ctgggtctcc	180
attctctaga	agaaatcgacc	tttaaaaggt	agaattaatt	tagttctcag	cgagagaactc	240
aaggaaactc	cacaaggagc	tcattttctt	tcagaagatc	tagatgatgc	cttaaaactt	300
actgaacaac	cagaattagc	aaataaaagta	gacatgggtc	ggatagttgg	tggcagttct	360
gtttataaag	aagccatgaa	tcacccaggc	catcttaaac	tattttgtgac	aaggatcatg	420
caagactttg	aaagtgcac	gttttttcca	gaaattgatt	tggagaaata	taaacctctg	480
ccagaatacc	cagggtgtct	ctctgatgtc	caggaggaga	aaggcattaa	gtacaaattt	540
gaagtatatg	agaagaatga	ttaa				564

<210> 4
 <211> 2158
 <212> DNA
 <213> Homo sapiens

<400> 4						
gcgcggcata	acgacccagg	tcgcggcgcg	gcggggcctg	agcgcgtggc	cggtgccgca	60
ggagccagcg	atggagtacc	aggatgcctg	gcgcgtgctc	aataccctgc	agaccaatgc	120
cggtcagctg	gagcaggtga	agcggcagcg	gggtgacctc	cagacacagt	tggaaacctg	180
ggaactgtac	ctggcaccga	gtgggctgca	ggtggaggac	ttggaccggc	tgaacatcat	240
ccacgtcaat	gggacgagcg	ggaagggttc	cacctgtgcc	ttcacggaa	gtatctcccg	300
aagctatggc	ctgaagacgg	gattctttag	ctctcccccac	gtggcgagg	ttcgggaagc	360
gatccgcgat	aatgggcagc	coatcagtc	tgagctcttc	accaagtagt	ctctggcgct	420
ctaccaccgg	ctggagagaga	ccaaggatgg	cagctgtgtc	tcctggcccc	cctacttccg	480
cttctcgaca	ctcatggcct	tccacgtctt	ccctcaagag	aagggtggac	tggcagtggt	540
ggagctgggc	attggcgggg	cttatgactg	caccaacatc	atcaggaaac	ctgtggtgtg	600
cgaggtctcc	ctctcttgga	tgcaccacac	cagcctccctg	ggggatacgg	tggagaagat	660
cgcatggccg	aaagggggca	tccttaagca	aggtgtccct	gccttcaactg	tgctccaacc	720
tgaaggtccc	ctggcgagtc	tgaagggaacc	agcccagcag	atctcatgtc	ctctataact	780
gtgtccgatg	ctggagggcc	tgcagggaag	ggggccggcg	ctgaccctcg	gcttgcaggg	840
ggagcaccag	cggtccaacg	ccgccttggc	cttgcagctg	gcccactgct	ggctgcagcg	900

gcaggaccgc	catggtgctg	gggagccaaa	ggcatccagg	ccagggtctc	tgtggcagct	960
gccccctggca	cctgtgttcc	agccccacat	ccacatgcgg	ctcgggtctc	ggaaacacgga	1020
gtggccgggg	cgagcccgagg	tgtctgcggcg	cgggccccc	acctggatcc	tggacgggtgc	1080
gcacaccgcc	agcagcgccg	aggcctgcgt	gcgctgggtc	cgccaggcgc	tgcagggccg	1140
cgagagcccg	agcggttgcg	cagaggttcg	agtcttgcct	tccaatgcta	ccggggaccg	1200
ggaccggcg	gccctgtctga	agctgctgca	gccccgccag	tttgactatg	ccgtctctctg	1260
ccctaaccctg	acagaggtgt	catccacagg	caacgcagac	caacagaact	tcaacgtgac	1320
actggaccag	gtcctgtctcc	gctgcctgga	acacaccagc	cactgggaac	acctggacga	1380
agagcaggcc	agcccgagacc	tctggagtgc	ccccagccca	gagcccggtg	ggctccgcatc	1440
ctgtctctctg	gcgcccacac	caccccacac	ctgcagtgc	agctccctgc	tcttcagctg	1500
catcttccat	gccttgcgaat	ggatcagcca	aggcccgagc	cccatctctc	agccaccctag	1560
tccccaaaag	ggcctctctca	cccaccctgt	ggctcacagt	ggggccagca	tactccgtga	1620
ggctgctgcc	atccatgtgc	tagtcaactg	cagcctgcac	ctggtgggtg	gtgtcctgaa	1680
ctgtctggag	cccgcaactgt	ccagtagcc	aaggcccggtg	gttgagggtg	ggagcttccc	1740
acacctgcct	gcgttctccc	catgaactta	catactaggt	gccttttgtt	tttggtcttc	1800
ctggttctgt	ctagactggc	ctagggggcca	gggctttggg	atgggaggcc	gggagaggat	1860
gtctttttta	aggctctgtg	ccttggtctc	tccttctctc	tggtgagat	agcagagggg	1920
ctccccgggt	ctctcaactg	tgcagtggcc	tggccgttca	gctgtctccc	cccaacacc	1980
cgctctgcct	ctggctcagg	ccagcttat	tgtgtgcgct	gcctggcgag	gcctcggttc	2040
tgcctatgtg	ctgggtggta	gatttctctc	tccagtgcc	ttctgggaag	ggagagggcc	2100
ctgctctggg	acactgcggg	acagagggtg	gctggagtga	attaaagcct	ttgttttt	2158

<210> 5

<211> 7720

<212> DNA

<213> Homo sapiens

<400> 5

taagtgtgaca	cttctcaggt	tgtcacaa	ttcaggtagt	gctcactgtt	gcaggacata	60
agctggggaat	tcttgggaat	tgtctgtctt	gcaggcccta	gagagccttc	cttcttggtt	120
gattttctctc	tagagatcca	actgtctctc	caggctcccc	tgctgccttc	ctccttgggt	180
ctttctctgt	ggcattggca	gattactggg	ccccattttt	ccctacactt	actgccactc	240
atagtctgat	ggttcccaca	tctgcatcca	acctggactc	ttccccctag	ctttcccctc	300
tacaaccgat	ttccccgggc	caagggcaca	caggcacctc	gacaaaacag	tgttctatgt	360
ttcttctctg	ccaaacctgc	ccctccctct	cccttttccc	atctgtggta	ccaccatggg	420
ctcagagaa	aaaaaaaaatg	aaggcttctg	tcattgaactg	gggtggagat	ggaggggaaga	480
gttagcccg	aatcacaggt	gctgtagaaa	ggatacctga	gttgccggga	gaggggggtcc	540
atgagttggg	gatgggaagg	gagcttggcc	cttcaaaaca	ttgaagatct	gatcaaaaga	600
ttcagaacat	ctgtgatttt	tggctgtgtg	atgggtgaca	ctggggtctaa	tggggttggg	660
ggagtttggtg	gctctacaat	tattggcctt	gggagatcct	tgctctctat	agctgactgg	720
gaggtttgaa	gcctctctag	tagcccttgc	cttgatcctc	cggatctcat	ttctctcact	780
tgcttaacag	gacagagggg	tggaaactg	atgagattag	ctcaaaagat	ctgggcagct	840
caggtctcaa	gatttttttc	agacctcagt	gtttgggaaa	aaattgggta	gggtgagctt	900
agggactggc	cttaggcctg	cactgttaat	tcacccccct	ccactcccc	atggaggcct	960
ggctgggtgc	cacatacaat	aattaactgc	tgagtggcct	tcggccaact	ccaggctcca	1020
ctcctgggct	ccattcccac	tcctgcctg	tctcttaggc	cactaaacca	cagctgtccc	1080
ctggaataag	gcaaggggga	gtgtagagca	gagcagaagc	cttagccaca	ccgagagcca	1140
ctctctctcc	caggttatgt	acactcccca	lcccccttca	gaggccacac	acctatggc	1200
attccccaca	tgtgttaagg	atttttctgaa	ctgggaaggcc	ctctgtgttg	ctgtaaggcc	1260
agagaattctt	gaagttggaga	ctgaggccca	gaccagagtg	tggcctgtct	aagatttaac	1320
gcacagattct	tgttctctcc	ctgtaactag	tacctgggct	ctagcccttc	agtcacagac	1380
tgaattctca	gctctctctag	tctggggccc	caaggttggg	tgggggttgc	atgattgtgtg	1440
gtgggggggg	gtcacagctg	gactaagacc	tgaaggttag	actaggcagg	tgggaaagga	1500
gcttccagag	gtatgtctgt	caaaaggaca	ggaagagagc	ctggcttcag	aagcagccac	1560
agcaagagag	actactgact	gaacaggtgg	gctccactgt	gggctccgga	aagatttttc	1620
tcagcccoca	tcccagcac	tgtgtgtgtg	cgcaccccat	gagagcctca	gcactcttga	1680
gggtcagggg	gcaaaggcca	aaagagctct	ggcctgaact	tgggtgtgct	ctactgtgtg	1740

acttggggca	tggccctcat	ctgtgctgaa	atgatccac	aaagattaaa	ctggctatca	1800
tttgttgatt	tcccctctct	tacatttaat	ccttgacaga	gaaagctaa	cctcaagata	1860
gtttgtctct	cttcccctca	aggccaagga	gaaggtggag	tgagggctcg	gagctgggaca	1920
ggttgaaacg	gaacccctgt	ctctaaacag	ttagggtttg	ttcccgcagg	aactgaacc	1980
aaagatcac	ctgttatctc	ctgagagatc	agatttctcc	ggcgtggccc	ctaaaggttag	2040
tgagtgaaga	ggctccacagg	ggcatgattg	gatccctgaa	tgaatgaate	aaccatgaga	2100
gagtgaatga	acaactggaat	caatagagta	gcagagtaat	ggaattgtgga	gcaggaaaag	2160
gagctgctgg	gtgggaattc	aattccaggc	ttatatgagc	cctgctgtgc	agtcggcctg	2220
gagacagccc	agctcaggcc	ctgcctagac	ccctgtcaag	gaggccctgt	caaggaggaga	2280
ggagggggcg	cacgggggca	aggcaagctt	gtgagcgga	aaggcatgtc	cacttagcg	2340
actggtatgt	ggaagatgag	ttagaggaga	cagatggaga	gaaagtcatg	gaaataaatt	2400
ctgagcatctt	taggagggcc	cagacacctg	gtgtccagt	gagtgaagga	aacagctgcc	2460
tcaccaaat	cagtgtctga	ggctaaagga	ttgaagtctt	gtgatgacca	aggagaagcc	2520
agctctgtgg	tagggggcac	aggagctccc	caaggcccca	ggcgtgtcca	gctgctgttc	2580
ccctgccacc	accatgttcc	tgtgacccca	ccccaccaag	atccatgggt	ttccggggaag	2640
ggcctactaa	actagcttga	gtgatgagc	tagaaaaggg	ctgggaccaa	ggtttaaaaa	2700
gcaaaaacaaa	ctaaccaaaa	ccacactgca	gcccccccaa	ctaaaaacatt	tttataaaact	2760
tttttttttt	ttttgagatg	gagctctcgt	ctgtcaccca	ggctagatgt	caatggcaca	2820
atcttggctc	actgttaact	ccacctctgt	gattcaagtg	atctctctgc	ctcagctctc	2880
cacgtagctg	ggactacagg	cacacgacac	cgacccacag	tcattttgtta	tttttagtag	2940
agacaggggt	tcactatgtt	ggccaggctg	gtctcaaaact	ctgacacctc	gggtgatccac	3000
ccacctcagc	cttccaaagt	ctgtgggatta	caggcatgag	ccaccgcgcc	cagcccatatt	3060
ttgtaaacctt	ttacaactgaa	gtaattttgt	gtcaaaatct	tgatgagaaa	ttaatgtgtag	3120
tttatgtata	gttttaattt	atccacctag	tgtaaactgtt	tcaccccgaa	atatacattt	3180
gattatgtgg	tatatgaaaa	aaatattttc	tttgtaatcac	ctttgatgaa	atctcaaaaa	3240
attttaacc	tgaaaacatt	gaataagcca	ttgtggacct	atggcaaaact	ctgggctatt	3300
tctgcatttt	gcccacattc	atccttgaat	tatatcact	gcaactctgt	accactctgga	3360
gaaggcaatg	aggctcaagc	cagggaaggg	ttgtgtctaa	tcctaccttt	cattggatct	3420
ggggaagctg	agggagatgt	gggcagggtc	ctatctgccc	cagggctccg	tcaggccccc	3480
accctctctg	agccctgcac	acaacttaag	gccccacctc	cgcattccct	gggtgccactg	3540
accacagctc	tttcttcagg	gacagacatg	gctcagcgga	tgacaacaca	gctgtcgtct	3600
cttctagtgt	gggtggctgt	agtaggggag	gctcagacaa	ggattgtcat	ggccaggact	3660
gagcttctca	atgtctgcac	gaacgccaa	caccacaagg	aaaagccagg	ccccaggagc	3720
aagtgttcac	agcagggtgt	ccagggggtg	atctgggggtg	gtgagggaact	ggctcaggaa	3780
gaggaaacga	ggacatggaa	atgccaaaac	ccatgggcac	tggtgaactg	aaagtggagga	3840
gcccttcact	ttgcattaat	atgggtgact	tatttcagag	acactgtgcc	taattgtcgtt	3900
acaatgccaa	cagttcacct	tcttgggtgt	tgagtttccg	cattacagaa	ataagggaagc	3960
aggcccaaat	gagagctctg	gaaatgaagt	tgagtgacc	catctctggg	ttgtctgatt	4020
tagggattta	gactgggaat	gactctctca	aagatctgag	ggaagaaact	gcacactgtg	4080
catagtggtc	tcctttctgc	cagccctaaa	cagctcaaga	agggaagagt	ttctcacatta	4140
tgaggtctgt	tgcaaaagcat	tttctctgag	acaaagtctc	acaaagtctc	catatgttgc	4200
ccaggctggt	ctcaaaattcc	tgagctcaag	acctcagccc	tcaccaaggt	ctcccaaggt	4260
tggaattaca	gaaatgcagg	gtacggccct	ctgaagcatc	ttgggtctat	catctcgcaa	4320
aactttgggc	tgtgtctctc	gaccacattg	gacctgaggt	ctccctataa	catttatttt	4380
gctcacacc	ctttaatatc	ctgaacatga	tgatataact	aaagaaaaag	cagagtgaaaa	4440
gtaattttgt	ggccagggtg	tacgggtcac	gcctgtaact	ccaacactgt	gggatgtcga	4500
gatgggcaga	tcacttgagc	tcaggagttc	gagaccagcc	tggggcaagt	ggcaaaaacc	4560
catctctact	aaaaaataaa	aaaaattagt	cagggtgtgt	ggacatgtcc	tcagctcccc	4620
gctactactg	agggtcagggt	ggcgagggtc	gttgagccca	ggaggcagag	attgtagatc	4680
gtgcactcgc	actccagcct	gggcaacaga	gtgagacctt	gtcaaaaagaa	agaaaagaacg	4740
aaaaaaagaa	agaaaaggag	gaagggaagg	gagggaaggaa	agggaaggag	gaaaaggagg	4800
gaggaagaaa	agggaaggaa	gggagagaaa	cttgtaatc	gcatttcttt	ttttttttct	4860
tgacatagag	ttttgtctct	gtgtcccgag	gtggatggca	gtggcacaat	ctcagctcac	4920
tgcacactcc	accctccagg	ttcaagtgat	tctcctcgct	cagcctctctg	agtagggcaa	4980
gcacaccaca	cccgacttaat	tttttggttt	ttgtttgttt	ttgtttgttt	gtatttttag	5040
tagagatggg	gtgtttcacca	gtgtgtccag	ctgtgtctgc	aactcctcac	ctcataactc	5100
gccccctctg	gcttcccaaa	gtgtctgagat	tacaggtgtg	agccactcgc	ccccggccta	5160

```

agtgacacatt tttttttttt attttttttt ttattttatt agatggagtc ttgctctgtt 5220
gccacagctt gaggtcagtg gcacaattct agctcactgc aaectccacc tcccagggtt 5280
aagcaattct tctgccttgg cctccagagt agctgggagc ataggcaccct gccaccatgc 5340
ctagctaatt tttgtatttt tagtagaaat ggggttttgc catgttggtcc aggtcggtct 5400
ccattcttga ccttaagtga tctgtccacc tccactcccc aaagtgtctgg gttaccaggc 5460
actatgtgag ccactgtgccc ggccccacatt ttaattatta gcttgtcagc ctttaagtaat 5520
gagattccagg aagcttgagg ataggcacac agggacatag ttccaagtgt tccgtaattt 5580
tgcagccatc acaagtttagt ttttaaggaa aaagatttagt tccataagttg tttctcaata 5640
acttataata aaataacatc cacaattgat tggctatata ttgttttttt gtatcacaaa 5700
ttccaaaaac agataaatggg tgaggcagct agtcaaggagc aaaacacttc ccaagtagct 5760
gggatctacag tgtgtccgcca ccacacttgg ctagtttttt gttttgtttat tttttgagat 5820
ggagcttctgc ctgtctgccc aggtctggagt cagtggtcat gatctcggtc cactcgaaac 5880
tccacctgcc ggggttcacac cattctcctg cctcagcctc ccaagtagct gggactacag 5940
gtgccagcca ccacgcccgg ctaatttttt gtatttttag tagagacggg gtttcacat 6000
tgttgccagg atgttcttga tctcttagcc tcgtgatcca cccgctcggg cctcccaaaa 6060
tgctgggagt acaggcgtga gccaccgcac ccggcctaatt ttttatattt tttagtagaga 6120
cggggtttca ccatgttggc caggctgggt tcaactctt gatctcaggt gatccacctg 6180
ccttggtccc ccaaagtgtc gggattacac aagtaagcca ctgcaccacg cctgggggtta 6240
caatttaaat tgccttttta ccttcaaatc tttgacacct cagtgaggct taactgtacc 6300
gcactattac actacaagtc cccatccgtc tctgtttaat tttgttccaa agcaaaaaatc 6360
agggtatgtg ttcattgttg taacccagtg ttctacaaaa gtacctgggt gagtagtaat 6420
aggatctcaa taaagcttga attaacaatt tttgtaatga ctgcaactcc agcaggagact 6480
ccctttctgccc ctcacactgt ccttgacggc cctctccctc aaagaggtcc caatagcaag 6540
tattttctct ggtgacttcc agtgggctgg ggaatcaagg actaagagggg gagacactgc 6600
atgtggaaata tctgtgctgt gctggctgtg ctggctgtgg actgactcct ctgttctccc 6660
ccatccagtg tgcaccttgg aggaagaatg cctgctgttc taccacacac agccaggaa 6720
cccataagga tgtttctcac ctatatagat tcaactggaa ccactgtgga gagatggcac 6780
ctgcctgcaa acggcatttc atccaggaca cctgcctcta cgagtgtctc cccaacttg 6840
ggcccttgat ccagcaggta tgcattggct cctgcaggta caagacctag cggagcagct 6900
gagcttttcca ggcactctct caggctgcaa ccccagctcc agttctatc gggctcagat 6960
tgctgggagt cctgaacctg agcccttctt ttgtatcaaa atcaccagg tggatcagag 7020
ctggcgcaaa gagcgggtgc tgaacgtgcc cctgtgcaaa gaggactgtg agcaatgggt 7080
ggaagattgt gcacactctc acactcgcaa gagcaactgg cacaagggct ggaactggac 7140
ttcaggtagg ggcctggggtg ggcaggaatg gagggatttg gaagtggagg ttgtggggtg 7200
tggaacaggt atgtgacaa ttggagttgt agggctggca gacctcaaga tagttccggg 7260
ccagttggct aaaggtcttc cctctctct acagggttta acaagtgcgc agtgggagct 7320
gcttgccaac ctttccattt ctacttcccc acaccactg ttctgtgcaa tgaaatctg 7380
actactcctt caaaggtcag caactacagc cgagggagtg gccgctgcat ccagatgtgg 7440
ttcgaccagc ccacaggcaa cccaatgag gaggtggcga ggttctatgc tgcagccatg 7500
agtggggctg ggcctctggc agcctggcct ttctctgcta gccctggcct aatgtctgct 7560
tggctgctca gctgacctcc ttttaccttc tgatactctg gaatccctc cctgttccag 7620
ccacagctgc ccaactattt ggttctctc ccatggtcgg gcctctgaca gccactttga 7680
ataaacaga caccgcacat gtgtcttgg aattatttgg 7720

```

<210> 6

<211> 255

<212> PRT

<213> Homo sapiens

<400> 6

Met Val Trp Lys Trp Met Pro Leu Leu Leu Leu Val Cys Val Ala

1

5

10

15

Thr Met Cys Ser Ala Gln Asp Arg Thr Asp Leu Leu Asn Val Cys Met

20

25

30

Asp Ala Lys His His Lys Thr Lys Pro Gly Pro Glu Asp Lys Leu His

35

40

45

Asp Gln Cys Ser Pro Trp Lys Lys Asn Ala Cys Cys Thr Ala Ser Thr
50 55 60

Ser Gln Glu Leu His Lys Asp Thr Ser Arg Leu Tyr Asn Phe Asn Trp
65 70 75 80

Asp His Cys Gly Lys Met Glu Pro Ala Cys Lys Arg His Phe Ile Gln
85 90 95

Asp Thr Cys Leu Tyr Glu Cys Ser Pro Asn Leu Gly Pro Trp Ile Gln
100 105 110

Gln Val Asn Gln Thr Trp Arg Lys Glu Arg Phe Leu Asp Val Pro Leu
115 120 125

Cys Lys Glu Asp Cys Gln Arg Trp Trp Glu Asp Cys His Thr Ser His
130 135 140

Thr Cys Lys Ser Asn Trp His Arg Gly Trp Asp Trp Thr Ser Gly Val
145 150 155 160

Asn Lys Cys Pro Ala Gly Ala Leu Cys Arg Thr Phe Glu Ser Tyr Phe
165 170 175

Pro Thr Pro Ala Ala Leu Cys Glu Gly Leu Trp Ser His Ser Tyr Lys
180 185 190

Val Ser Asn Tyr Ser Arg Gly Ser Gly Arg Cys Ile Gln Met Trp Phe
195 200 205

Asp Ser Ala Gln Gly Asn Pro Asn Glu Glu Val Ala Arg Phe Tyr Ala
210 215 220

Ala Ala Met His Val Asn Ala Gly Glu Met Leu His Gly Thr Gly Gly
225 230 235 240

Leu Leu Leu Ser Leu Ala Leu Met Leu Gln Leu Trp Leu Leu Gly
245 250 255

<210> 7

<211> 817

<212> DNA

<213> Homo sapiens

<400> 7

cgcaggaata gatggacatg gacctggcaga tgaatgcagct gctgcttctg gctttgggtga 60
ctgctgcggg gagtgcacag cccaggagtg cgcgggcccag gacggacctg ctcaatgtct 120
gcatgaacgc caagcaccac aagacacagc ccagccccga ggacgagctg tatggccagt 180
gcagtccctg gaagaagaat gcttgctgca cggccagcac cagccaggag ctgcacaagg 240
acacctcccg cctgtacaac tttaactggg atcactgtgg taagatggaa cccacctgca 300
agcgccaatt tatccaggac agctgtctct gagtgtctac ccaacctggg gccctggatc 360
cggcagggtca accagagctg gcgcaaaagag cgcattctga acgtgccct gtgcaaaagag 420
gactgtgagc gctggtggga ggactgtcgc acctcttaca cctgcaaaaag caactggcac 480
aaagctgga attggacctc agggattaat gagtgtccgg cccggggccct ctgcagcacc 540

tttgagtctc	acttccccac	tccagccgcc	cttctgtgaag	gcctctggag	ccactccttc	600
aagggtcagca	actatagtctg	aggagccggc	cgctgcatcc	agatgtgtgtt	tgactcaagcc	660
cagggtcaacc	ccaattgagga	ggttgcccaag	ttcttatgctg	cgcccatgaa	tgtctggggccc	720
cgctctcgtg	ggattatttga	ttcctgatcc	aagaagggtc	ctctgggggtt	cttccaacaa	780
cttattcttaa	tagacaaatc	cacatgaaaa	aaaaaaa			817

<210> 8
 <211> 1669
 <212> DNA
 <213> Homo sapiens

<400> 8						
gctaggcagc	ttcgaaccag	tgcaatgacg	atgccagtca	acggggccca	caaggatgct	60
gacctgtggg	cctcacatga	caagatgctg	gcacaacccc	tcaaagacag	tgatgtttgag	120
gtttacaaca	tcattaagaa	ggagagtaac	cgccagaggg	ttggattgga	gctgattgcc	180
tcggagaatt	tcgccagccg	agcagttttg	gaggccctag	gctcttgctt	aaataacaaa	240
tactctgagg	ggtaccgccg	ccagagatac	tatggccgga	ctgagtttat	tgatgaactg	300
gagaccctct	gtcagaagcg	agccctgcag	gcctataagc	tggaaccaca	gtgctggggg	360
gtcaacgttc	agccctacac	aggtctccct	gcaaaccttg	ctgtgtatcac	tgccctgggt	420
gaaccccatc	ggcgcatcat	gggcctggac	cttcgggatg	ggggccacct	gaccocattgg	480
ttcatgacag	acaagaagaa	aatctctgcc	acgtccatct	tctttgaatc	tatgccctac	540
aagggtgaacc	caagatactgg	ctacatcaac	tatgaccagc	tgaggaggaa	cgacgcgctc	600
ttccaccgca	agctgacatc	cgcaggaacc	agctgctact	cccgaacctt	ggaatatgcc	660
cggtctacgga	agattgcaga	tgagaacggg	gcgtatctca	tgccgggacat	ggctcacatc	720
agcgggctgg	tgccggctgg	cggtgtgcc	tcccatttg	aacactgcca	tgtgtgtacc	780
accaccactc	acaagacctt	gcgaggtctg	cgagctggca	tgatcttcta	caggaagaaga	840
gtgaaaagtg	tggtatccaa	gactggcaca	gagattctgt	acaacctgga	gtctctttatc	900
aatctctgct	tgttccctgg	cctgcaggga	ggtccccaca	accacgccat	tgctgggggt	960
gctgtggcac	tgaagcaagc	tatgactctg	gaatttaaa	gttatcaaca	ccaggtgggt	1020
gccaaactgc	gggactctgt	tgaggccctg	acggagctgg	gtacacaaat	agtcacaggt	1080
ggttctgaca	accatttgat	ccttgtggat	ctcgtttcca	aaggcacaga	tggttgaagg	1140
gctgagaagg	tgctagaagc	ctgttctatt	gcttgcaaca	agaacacctg	tccaggtgac	1200
agaagcgctc	tcgcggccac	tggaactcgg	ctggggaccc	cagcactgac	gtcccggtgga	1260
cttttggaaa	aagacttcca	aaaagttagc	cactttattc	acagagggat	agagctgacc	1320
ctgcagatcc	agagcgacac	tggtgtgcaga	gccaccctga	aagaggtcaa	ggagagactg	1380
gcaggggata	agttaccagc	ggccgtgcag	gctctccggg	aggaggttga	gagctctgcc	1440
tctctcttcc	ctctgctctg	cctgcctgac	ttctaaagga	gcggggccac	tctggaccaca	1500
ctctggcgcca	cagaggaagc	tgccctgcgg	agaccccac	ctgagagatg	gatgagctgc	1560
tccaaaggga	actgttgaca	ctcgggccc	ttgaggggtt	ttcttttggg	cttttttcat	1620
gttttcttca	caaatcaaaa	ttgttttaag	tctcatgtgt	agtaattctt		1669

<210> 9
 <211> 3112
 <212> DNA
 <213> Homo sapiens

<400> 9						
gtggaacctc	gatattgggt	gtgtccatcg	tgggcagcgg	actaataaag	gccatggcgc	60
cagcagaagt	cctgaacggg	aaggagatct	ccgcgcgaat	aaggggcgaga	ctgaaaaaatc	120
aagtcaactca	gttggaacgg	caagtacctg	gtttcacacc	acgcctggca	atattacagg	180
ttggccaacac	agatgattcc	aatctttata	taaatgtgaa	gctgaaggct	gctgaagaga	240
ttgggatcaaa	agccactctc	attaagtac	caagaacacac	cacagaatct	gctgtgatga	300
agtaacattac	atctttgaat	gaagactcta	ctgtacatgg	gttcttagtg	cagctacatt	360
tagattcaga	gaatttcatt	aacactgaag	aagtgtatcaa	tgctattgca	cccgagaagg	420
attgtggatgt	attgactagc	atcaatgctg	ggagacttgc	tagaggtgac	ctcaatgact	480
ttgtactctc	ttgtactgct	aaggatgctt	tggaaactcat	caaaagacac	gggggtgcga	540
ttgccggaag	gcattgctgtg	gtggtttggc	gcagtaaaat	agttggggcc	ccgatgcatg	600

acttgcttct	gtggaacaat	gccacagtga	ccacctgcca	ctccaagact	gcccactctg	660
atgaggaggt	aaataaagg	gacatcctg	tggttgcaac	tggtcagcct	gaaatcggtg	720
aaggggaggtg	gatcaaacct	ggggcaatag	tcactcgactg	tggaatcaat	tatgtcccag	780
atgataaaaa	accaaatcgg	agaaaagtgt	tggttgatgt	ggcatcacgac	gaggccaaag	840
agaggggcag	cttcatcact	cctgttctctg	gcggcgtagg	gcccattgaca	gttgcaatgc	900
tcactgcagag	cacagtgaac	agtgccaagc	gtttcctcgtg	gaaattttaa	ccaggaaagt	960
ggatgtattca	gtataacaac	cttaacctca	agacacctgt	tccaagtgcac	attgatatat	1020
cacgatcttg	taaacccga	cccatttgta	agctggctcg	agaatttggt	ctgctgtctg	1080
aagaggtaga	attatatggt	gaaacaaagg	ccaaagtctct	gctgtcagca	ctagaccgcc	1140
tgagaccacg	cgcctgactg	aaatacgtgt	tggtgactgg	aataactcca	acacccctgg	1200
gagaaggagaa	aagcacacat	acaatcggcg	tagtgcaagc	ccttggtgccc	catctctacc	1260
agaatgtcttt	tgctgttggt	cgacagcctt	ctcaggccccc	cacctttggga	ataaaagggt	1320
gcgctgcagg	agggcggtac	tcccaggtca	ttcctatgga	agagtttaat	ctccacctca	1380
cagggtgacat	cactgcacatc	actgcagcta	ataacctcgt	tgctgcggcc	attgatgctc	1440
ggatatttca	tgaactgacc	cagacagaca	aggctctctt	taactcgttg	gtgccatcag	1500
taaatggagt	gagaagggtc	tctgacatcc	aaatccgaag	gttaagaga	ctaggcatcg	1560
aaaagactga	ccctaccaca	ctgacagatg	aagagataaa	cagatttgca	agattggaca	1620
ttgatccaga	aaccataact	tggcaaaag	tggtggatag	caatgataga	ttcctgagga	1680
agatcacgat	tggacaggct	ccaacggaga	agggctcacac	acggacggcc	cagtttgata	1740
ctctctgtggc	cagtgaat	atggctgtcc	tggtctctcac	cactctctca	gaagacatga	1800
gagagagagc	gggcaaaatg	gtggtggcat	ccagtaagaa	aggagagccc	gcagtgccg	1860
aaagtctcggg	gggtgagctg	gcactgacag	tgcttatgaa	ggacgcaatc	aagcccaact	1920
tcactgcagac	actggaggctc	actccagtg	ttgtccatgc	tgcccgcttt	gccaaactcg	1980
cacatggccaa	ttctccatc	attgcagacc	ggatgcgact	caagctgtgt	ggcccagaag	2040
ggtttgtagt	gacggcaatga	ggatttggag	cagacattgc	agtggaaga	tttttaaca	2100
tcaaatcccg	gtattccggc	ctctgccccc	acgtggtggt	gcttgtgtcc	actgtcagg	2160
ctctcaagat	gcacgggggc	ggcccccacg	tcaactgctg	actgcctctt	ccccaggctt	2220
acatacagga	gaacctggag	ctggttgaaa	aaggcttcag	tacttgaa	aaacaaattg	2280
aaaatgccag	aatgtttcgt	attccagtgt	tagtggccgt	gaatgcattc	aagacggata	2340
cagagtctga	gctggacctc	atcacgcgcc	tttccagaga	acatggggcc	tttgatgccg	2400
tgaagtgcac	tcaactgggca	gaagggggca	agggtgctt	agccctggct	caggccgtcc	2460
agagagcagc	acaagcacc	agcagcttcc	agctcttcta	tgacctcaag	ctccagtg	2520
aggataaaaa	caggatcat	gcacagaaga	tctatggagc	agatgacatt	gaattacttc	2580
ccgaagctca	acacaaagc	gaagtctaca	cgaagcagg	ctttgggaat	ctcccatctc	2640
gcatggctaa	aacacactgt	tctttgtctc	acaaccagga	gcaaaaagg	gtccctacag	2700
gcttcattct	gccattcgc	gacatccgcg	ccagcgttgg	ggctggtttt	ctgacctcct	2760
tagtaggaac	gatgagcaca	atgcctggac	tccccaccg	gcctgtgttt	tatgatattg	2820
attttggacc	tgaacagaa	caggtgaatg	gattattcta	aacagatcac	catccatctt	2880
caagaagtca	ctttgaaagt	ctggccagtg	tctattcagg	ccactggga	gttaggaagt	2940
ataagttaag	caagagaagt	cagccctgc	ccagaagatc	tgaactaat	agtaggaagt	3000
tcccacgaag	tcattttgag	ccttaattct	catcatgtat	aaattacat	ataatcatgca	3060
tgtctgttta	ctttagtgac	gttccacaga	ataaaaggaa	acagtttgcc	ca	3112

<210> 10
 <211> 1792
 <212> DNA
 <213> Homo sapiens

<400> 10						
cgcagcccgac	actcagactg	gggaagcaaa	caggggcttg	acaggccagg	agagcctgtc	60
ggacagtgat	cctgagatgt	gggagttgtc	gcagagggag	aaggacagcc	agtgctcgtg	120
cctgagcgtc	attgcctcag	agaactctctg	cagcccgagct	gcgctggagg	ccctggggct	180
ctgtctgaac	aaacagtcact	cggagggtta	tcttggcaag	agatactagt	ggggagcaga	240
gggtgtggat	gaaattgagc	tgctgtgcca	gcgcggggcc	ttggaaacct	ttgaccttga	300
tctctgcagac	tggggagtgca	atgtccagcc	ctactccggg	tcccagacca	acctggccgt	360
ctacacagcc	ctctctcaac	ctcagcagcg	gatcatgggg	ctccagctgc	ccgatggggg	420
ccatctcacc	cacggctaca	tgtctgacgt	caagcggata	tcagccacctg	ccatcttctt	480

aagaggacag	actgaagct	ggcggggaa	cttggtttcc	tgttggtcct	ccatccaatc	1740
ccacgaacac	agcttttctc	ttaaaccttg	aaaagagaaa	tctgggaggt	cgagttctta	1800
ctcgtctctt	ctctcttctc	tcccgacagg	agcacccacg	gcacaaatag	tctcggcggt	1860
catttggtccc	agggctttcag	gggacagtg	ggcggggcgg	gggtgggcaca	ggacgttagg	1920
cagccgttgg	ccctccctaa	ggccacacgg	tctcgccgtc	ctggatctctg	gccacgtcgc	1980
cggggggagg	ggactcgaag	gtgtgtgagc	caggggctga	ccttgaccgc	tcagataaat	2040
ggagccgacg	cttgacacag	gggtggagg	ggttttgaat	ggggaaaccc	attcgttggtg	2100
aagcagattc	actgtagcta	gcggaaaagc	cctccggccc	acggaccatc	ctagagacga	2160
atcatataga	gcgtgctgtg	ctgattggcg	tgggacagcg	tggtggagtt	tgtctgagga	2220
gagggatcca	cttttctgca	gctccaaagc	caggggccct	tgatgagcca	tagacctcat	2280
tttttaaccca	cttttctgct	tagacattga	gcaagttact	tctcatatag	cttccctata	2340
tgttaaaaat	ggagaaaaata	atgcttagta	ggcaattctg	ataaaatctg	gtgcttgcaa	2400
aaatctctct	gttgtctgaa	tataaacgtg	accacaagcg	agtcggcgatg	aacgaggact	2460
gcatttaaa	ataagttttt	acactttcat	tctctgtgtg	ctcgacactt	ctgatgcctc	2520
ctttttttgt	cctgggacac	atgcttggtg	ttgtcttcac	acottttgtga	caggattagc	2580
actagtgggc	agtggatgat	agctcctcct	ccottttgcc	acatgttcat	ccttgccctc	2640
gccaccatct	cactgtgtgg	aattcctgtg	tccactggtc	acggggggac	agaagtgtctg	2700
tctcagctg	aatcgggcca	ctgatgggac	ttgcagcctg	ggagctccac	cgtagctctc	2760
ggcccaactt	gcgggagctg	aggctttctg	gatgtccag	gcctcacgtg	ccagggcgagt	2820
tttttctccc	gaagaaaggt	ggatggcgat	atctgtcttc	ccattctgaa	accgtatggc	2880
aaatgtgttt	tcagatgaat	tccctctgct	gacacacaaa	ctgtgtgtct	ggagaggggtg	2940
tttgaggagg	ttgtctgtgt	ttatcaaggt	aaagaagtcg	ctgctattag	aagtcagtag	3000
tctgtttcca	acacagcagc	cagtgagatc	ctttcaaaac	tcaaagcagc	caggtgtgggt	3060
ggctcacgct	tgtaatccca	cgcgtttggg	aggctgagtc	agatcacctg	aggttaggaa	3120
tttgggacca	gccttggccaa	catggcgaca	cccagctctc	tactaataac	acaaaaaatt	3180
agccagggtg	gctggtgcac	gtctgtaac	ccagctactc	aggaggctga	ggcatgagaa	3240
ttgtctcaga	ggcggagggt	gtagtgaact	gagatcgtgg	cacgtgactc	cagcctggcg	3300
acagagggag	aacccctgtc	aaaaacaaaa	aaagacacca	ccaaaggctc	aagcatatca	3360
ttctccaccc	tcaagccctt	agtggctcca	tttccactcag	taagagccac	ggctcctatg	3420
gtgtccgttt	ttcagctgtg	accttagctg	ctgctctctg	caccacactc	gtctcttctg	3480
gagtttttga	gcacaccggg	acatccccc	tccttggaac	cttcttcccc	cacactttggc	3540
ttctctcttt	gagtctctac	tccactcggg	caagccttcc	tagacctctc	gattttaaacc	3600
tgtgactctc	ccccaacctc	cttggtgttt	ctccgtagac	gaacatcacc	atctgatgta	3660
tgctcagctt	tcctctcccc	tgttagaagg	gggacagcag	gtagtaaaag	tgaaaatgtgc	3720
tgtaagcttt	atgaggcgag	aggatttgtt	tctgtgttcc	actgtttgat	cgccaggggc	3780
tcaaacacag	cctgccacct	agtaggagtc	aacatatatt	gatcactaaa	tgtagatacc	3840
acctgtgttc	ccatgttcat	ataaatctca	gaagagtctc	ttcagtaaca	aggtgaaccc	3900
cttccagagg	gctgagtagg	tacctcagcg	cggggccaga	gtgctgtgaa	gacagcagca	3960
gcccagagcca	agcttctctg	tgttccgtgt	cctggtctag	aacaccagat	gtttcttctg	4020
accagtgctt	tttggaaagg	ggctgaggtc	tggggtcagg	cttggggcat	actagaagctg	4080
gggactccct	ctatagagca	cttggtatgg	cttgatgggt	cttggggcaa	gccagagccca	4140
agccctctta	tccattttta	gaaagggtct	caatttggat	ccagccccag	gtctgcctta	4200
gctctgtatt	cttggggtat	ttgtttctgt	attggcctat	cttgactaac	aatgagcctt	4260
gattttgaaa	catatcatca	gaaacctcag	aagacaacat	tcttaaacctg	gctagagcct	4320
ggcttgaagt	gatgaaaagg	agagactttt	gaagcaatat	gtaaaaggat	gagaaatgat	4380
ttgttggaag	tttctcaatt	ggagaaaatt	ctttgatttg	ttggaaaatt	ctttgatctc	4440
ttctcaatca	agaagaaatc	ggacaacatc	aacaatagaa	aggggaggag	caagatactc	4500
agaaaataaaa	tgcattcccc	tgtttcaact	taatgtctca	attcaggatt	ctaagggaatc	4560
cttgcaggaa	atgtcagact	caacctgata	gttggagtta	cttccattgtg	gactcgatca	4620
aatacaggag	tggaggcacc	tgcactgtaa	aatactgatt	atgctgatca	ttaggaagctc	4680
ctgtatggc	aggtagaaga	tacattgaac	agattgcatg	taggcattaa	attcattttg	4740
gggtattaca	tatagacaac	acattttcat	agaagacata	aaactgtcag	atcgggtgga	4800
tactttaaag	cacttggagg	gtgttagcct	aaaaagctta	gttgaggggca	atggaagaaa	4860
agatctggga	gggtgtttcc	aaagaaggga	tcagactatc	ctaaagccct	caggaatctg	4920
ggctggggac	acctacttaa	agataggatg	ggcagctggg	tgtgtgtgct	cacgcctgta	4980
atccagcac	ctcggggagg	cgaagcgggc	ggatcacctg	aggtcaggag	ttcggaggcca	5040
gcctgagacca	cattgagaaa	cgtctctctc	actaaaaata	aaatttgatg	tgttggtagt	5100

ggcgcatgcc	tgtaatccca	gctactcggg	aggctgaggg	aggggaatcg	cttgaacctg	5160
ggaggtggag	gggtgcgtga	gccacgatcg	cgccattgca	ctccagcctg	ggcaacaaga	5220
gcgaactctt	caaaaaacaa	aaaaaaggat	gggttcacata	tgggtgggtg	caagtgccca	5280
ctctctagca	agtcagcagg	ggccagaggg	ccttgtaagt	gggtctctcg	ggggatacaac	5340
tgagatggct	taagatttac	ctggatgcct	gctctgctct	ccccatctct	tcaggggatac	5400
cacaaatgct	aaagagctgt	cttccaaggg	agtgaaaaatc	tgggatgcga	atgggacctg	5460
agactttttg	gacagctcgg	gattctccac	cagagaagaa	ggggacttgg	ggccagttaa	5520
tggcttccag	tggaggcatt	ttggggcaga	atacagagat	atggaatcag	gtgaggagat	5580
agaacaatgc	cttccattct	cggttgccct	tcttagcacg	tgtttgctcc	gtgtgttttag	5640
ataaggtctg	ggggatgagt	caatgtcaca	ggagctgatg	tatagctttg	acctgtgtgag	5700
gggtgggtgc	aggttgtaagc	cacaattaac	gcctactgaa	ggcgttttca	catctttttt	5760
tttttttttt	ttttaattat	tatactttaa	gttttagggg	acatgtgcac	acctgtgcagg	5820
ttagttaacat	atgtatacat	gtgcaatgct	ggtgcgctgc	accactaaat	caccatctag	5880
catocaggat	atctoccaat	gctatccctc	ccccctctc	ccaccocaca	acatccccag	5940
agtggtgatg	tccccttctt	gtgtccatat	gttctcgttg	ttcgatctcc	acatagagtg	6000
agaatatgcg	gtgttttggt	ttttgtttct	gcgatagttt	actgagaatg	atgattttca	6060
tttccaccag	tccctacaga	ggacatgaac	tcatcatatt	ttagtggctg	atagatttcc	6120
atgggtgtata	tgtgccacat	tttcttaact	cagtctatca	tgttggacat	tgggtgtgtg	6180
tccaaagtct	tgcctattgt	gaatagtgc	acaataaaca	tacgtgtgca	tgtgtcttta	6240
tagcacatgc	atttaatatg	cctttgggtg	tataccagat	aatgggtgag	ctgggtcaaa	6300
tggtaattct	agttctagat	ccccaggaa	tcgccacact	gacttccaca	atgggtgaac	6360
tagttttacag	tcccaccaac	agtgctcaaa	tgtcctattt	ctccacatcc	ctccagcagc	6420
ctgtgttttc	ctgacttttt	aatgattgca	attctaaact	gtgtgagatg	gtatctcatt	6480
ttgtgttttg	tttgcgtttc	tctgatggcc	agtgatgggt	agcatttttt	catgtgtttt	6540
ttggctgcac	aaatgtcttc	ttttgagaag	tgtctgttca	tgtctttttg	acctcttttg	6600
atgggggtgt	ttttttctta	taaaatttgt	tgagttcatt	gtagattctg	gatattagcc	6660
cttttgcaga	tgataggtgt	gcaaaaaatg	tcttccattt	tgtgggttgc	ctgttctact	6720
tgatgttagt	tcttttgtct	gtgcagaagc	tcttttagtt	aatatagatc	catttgttaa	6780
tttttggctt	tgttgccatt	gcttttgcca	taggcgatga	gtcccttgcc	atgcctatgt	6840
cttgaatggg	aatgcctatt	ttttcttcta	gggtttttat	ggttttaggt	ctaacgttta	6900
agtccttaat	ccatcttgaa	ttagattttg	tataaggtgt	aagggaagga	tccagtttca	6960
gctttttaca	tatggctcag	cagttttccc	agcacatttt	attacatagg	gaatcttttc	7020
cccatctgct	gtttttctca	ggtttgcaca	agatcagata	gttgtagata	tgcggcgcta	7080
ttcttgaggg	ctctgttctg	ttccattgat	ctatgtgtct	gttttggtag	gaatccattca	7140
ctgttttggg	tactgtagcc	ttttagtata	gtttgaagtc	aggtagcgtg	atgcctccag	7200
ctttgttctt	tgtgcttagg	attgacttgg	cgatgcggcg	tcttttttgg	tcccatatga	7260
actttaaagt	agrttttttc	aatctctgta	agaaagtcat	tggtagctgt	atggggatgg	7320
catgtaactt	ataaaattacc	ttgggcagta	tggccatttt	cacgatattt	attcttctca	7380
cccatgagca	tggaaatggg	ttccatttct	ttgtatctct	ttttatttca	ttgagcagtg	7440
gtttgtagtt	ctctctgaag	aggtctctca	catccctttt	aaggtggatt	cttaggtatt	7500
tattctctct	tgaagcaatt	gtgagtggaa	gttccactcat	gatttggctc	gtcttttctc	7560
gtttatttgt	gtataaagaat	gcttgtgatt	tttcgagatt	gattttatat	ctctagacatt	7620
tgtggaagct	gctttacagc	ttaaggagat	tttgggctga	gacaaatggg	ttttctagat	7680
atacaatcat	ctctgtctga	aacagggaca	atttgacttc	ctcttttctc	aattgaaatac	7740
ctttttattc	ctctctctgc	ctaattgcct	tggccagaaac	ttccaacact	atgttggaata	7800
ggagtggtga	gagagggcat	ccctgtcttg	tgcacagttt	caaagggaaat	gcttccagtt	7860
tttgcacctt	cactatgata	tggctgtgtg	ctttgtcata	gatagctctt	attattttga	7920
aatatgtttc	atcaataact	aatctattga	gagtttttag	tattagttgt	tgttgaattt	7980
tgctcaaggc	tttttctgca	tctatttgaa	taatcatgtg	gtttttgtct	ttggaattct	8040
ttatattcgt	gattacattt	attgatttgc	gtatatgtaa	cagactctgc	ctctcaggga	8100
tgaagcccaac	atgatcatgg	tggataagct	ttttgatgtg	ctcgtggatt	cggtttgcca	8160
gtatttttct	gaggaatttt	gcatacaagt	tcaatcaagg	tattgtgtca	aaattctctt	8220
ttttgtgtgt	tctctgcccc	gctttgtgat	caggatgatg	tgtccttcat	aaaatgagtt	8280
aggggaagatt	ctctctcttt	ctattgattg	gaatagtttc	agaaggaagt	gtaccagttc	8340
ctctttgtcat	ctctggagaa	tctggcgtgt	aatccatctg	gtcctggact	ctcttttggtt	8400
ggtaagacat	tgtattttgc	cacaatttca	gctcctgtta	ttgtcttcta	cagagattca	8460
actttctctt	ggtttagtgt	tgggagagtg	tatgtgtcaa	ggaattttatc	cattttctctt	8520

agaggttgtaa	agaaccccg	cgtcttcatt	tatactaacc	atactcttag	aggggaagcaa	12000
tctggtttttg	tgcagagggca	ctgaggggagg	caggaccctg	ggcaactctag	cccagccaca	12060
tgggtgtgtgt	acgttggggca	agtcacattt	tgtctgacct	tcacctcttcag	atcatgaggt	12120
tgggcccgaca	ggattttttt	tttttttttt	tttttttgaga	cagagattctt	ctctgttgcc	12180
caggctggaa	tgcacgggct	tgtatcttggc	tcactgtaac	ctctgctcttc	tgggttcgag	12240
tgattcttct	gcctcagcct	ccaagttagct	gggattacag	catgtgccac	catgcctggc	12300
taattttgta	tttttagtag	agacggggttc	acatgttgg	caggctggct	tgcactctgt	12360
accctcagat	gatctgacct	gcctcagcct	cccaaccgag	tgatcttaag	ttgtgtatta	12420
tactcattct	tacacataaa	gggctttaaa	tgccatagaaa	ctacatgaag	atgtttaacat	12480
ttttaaaggga	agcagatgaa	gttcacagct	cgctgcacct	cactaacatt	tttaacaatt	12540
atattgttaaa	attcacaact	accagggtgt	agagccaggt	gtgggtggctc	acacctgtaa	12600
ttccacaacac	tccagaggcc	aaggcgagag	gatcatttga	acccacggaa	tttgagctgt	12660
tagtgagtca	tgtatccgcc	attgcactcc	atctctggga	acagagtgag	accctgaata	12720
tttaaaaaca	acaacaacaa	caaaaactcta	tcaggatata	ataagtagct	agagtgaat	12780
acttgcattct	gtaattagaga	cttatttttt	ttttttttga	gacacagctc	cacctgtgtg	12840
cccaggctgg	agtgacgtgg	tttgcattcc	gctcacggca	acctccatct	cccaggttca	12900
agtgagttcc	cattcctcag	ccccagagct	gggaccacag	gcgcgcgaat	ttttgtattt	12960
ttagcagaga	cgggggttca	ctatgttggc	caggctagtc	tcaaaactca	gttggctcca	13020
agtgatctgc	ccacctctgc	gtccagtg	tggttatcca	ggcatgagcc	actgtgctgt	13080
gccatgtaat	agagactttt	aatataggag	gggtgaccag	aagcaccagt	ttcctgtggc	13140
aaacagaatt	attcctgtctg	tatttgaat	ttggtgccac	gaggttagccc	agatcccttc	13200
agctctgtat	gaagagcatt	gcttcagccg	taaatggaca	cctgcagaaa	ccttgcaacc	13260
tttgatagtc	tccctcagct	cctgtccatc	gtctcagggg	ctgttatgga	catcaactga	13320
gcccaattgc	tctctctctc	ggctccacc	atatgagttg	gcttctgttt	ctctcctgtt	13380
ttactttggc	tttagctgtg	gtctttcaaa	ccacctcccc	tccttatctt	ctctgtctgtg	13440
ttcctcagat	cttctctcga	tggcgctgcc	tccatgccat	gcctctgccc	agttctatgt	13500
gggtgaacagt	gagctgtctc	gccagctgta	ccagagatcg	ggagacatcg	tcctcggtgt	13560
gcctttcaac	atgcgcagct	acgcctctgt	cacgtacatg	atttgcgcaca	gcccgggcct	13620
gaaggtgggc	tgtctcggga	agggtagctt	gccagccctac	cacatgagct	cttcagttct	13680
ttaatatggg	aaaacaaaat	gcagagttaa	gtctctgatt	acgttttaaa	tttgatatgt	13740
gtaagtgaaga	catgaaccag	ctttactctt	gaaaccttcc	ttttctggaa	ggttttctgtg	13800
ccctgtggga	tatgcactaa	cagatctata	caggttgttt	gtgatacacg	ttctatggat	13860
cttctcaaaa	gctatgctga	ggttgggtaa	gggtggctcat	gcctgtaatc	ccagcacctt	13920
ggaagactga	gacaggagca	attgcttgag	gtctggagtt	caataccagc	ctggggcaaca	13980
taacaagatg	ctgtttcctac	aaaaaaaaatg	aaaagctaca	ctaaattatt	tttttaaaaa	14040
aagccttgcg	gtgtctgcatt	attctaagt	ttttaaatga	tgttttaaa	aattgaaact	14100
aaacatctgt	tctgtctttc	cccggtttat	agccaggtga	ctttatacac	actttgggag	14160
atgcacatat	ttacttgaa	cacatcgagc	cactgaaaat	tcaggtaaga	attagatgtt	14220
atacttttgg	gtttgggtacc	ttctcttgat	aaaagggtga	ctgttggaact	ggatctgct	14280
caatgtcgta	tccaagataa	agatgactgc	tccaatgtg	gggcttcagt	ttaggggaga	14340
gtggtgggca	gggtggcagg	acaaggcagg	actctgcttc	agcaacctag	gcacttaact	14400
tgtcaggtgc	tgtgaggtac	taagcaccag	taccagagag	ggaagagcca	cattcaagcc	14460
aggggagtgt	ccaaaaggag	gcattttaac	tcattttaac	tgtgaaggta	attgaagtgc	14520
aaatgttttt	ctcttttttt	ttttttgaga	tggagttctt	ctctgtctgc	agggctggag	14580
tgtgcctgtg	tgcgtatcca	gctcactgca	acctccacct	cccggtttca	agcaattctt	14640
ctgcctcagc	ctccaggat	gctgggatta	caggccacat	ccaccacacc	cagctaaatt	14700
ttgtatttat	tagtagagat	ggggttctgt	catgttggcc	gaattctgct	caactcctgt	14760
acttcaagtg	taccacctgc	ctcagcctcc	gaaagtctgt	aggttatcag	cataagccac	14820
caccttgggc	ataaataatt	tttgttaatt	ttacatgaag	tcaaatattt	aggtcccaac	14880
ttcaaaagtc	tgttgaaatt	cttgaagtta	tagcagccaa	taattgatat	gaaatggcaa	14940
taaaaaatgta	agttcattct	cttcagagc	cttaaggaaa	aaaaactcaga	accagacact	15000
ttttagcccc	ttccagggtta	gatccagggt	ttaaaagtta	ttcctttggg	ttgtgtttggc	15060
tgtcttttgag	tggagggtgac	ttcagggtta	ttctctctgg	ctctctgctc	tgtgtcaatt	15120
tagacatagt	aaatggttgt	gacctgtctt	cacatccctaa	tgtccactac	ctgttcatcc	15180
caggaaatcct	ggctttctgt	cctttcttgt	cactgtccat	gcagtgcate	tttctctctt	15240
cttgcacagg	accagatgag	ttagggtattg	tgaattcaag	taaacgtgaa	gctactatga	15300
gttacagatt	gactgtgttc	ctgtctttaa	taaatgtgcc	aagagtgggt	ataagaactt	15360

acacctgatg	aggcaccagg	ctcctgatgc	tgtgtaatgt	cacaaaaatc	ccctcactct	15420
cgatctgtgc	aaggaacacg	ctggttgccg	tccaatcatg	ttacataacc	tacgcgaagg	15480
tatcgacagc	atcatactcc	tgtaaaaatg	aactttgttg	atcacatcct	gtgtacttgt	15540
ttcacggaca	tgaggagcaa	ttacaacagg	tcgtacaatt	atggcaaaaa	aatggccctt	15600
ttttgttttt	agcttcacgc	agaaccagga	cctttcccaa	agctcaggat	ttctcgaaaa	15660
gttgagaaaa	ttgatgactt	caaagctgaa	gactttcaga	ttgaagggtt	caatccgcct	15720
ccaactatta	aaatggaaat	ggctgtttag	gagctcttca	aaggagctcg	aaggatattg	15780
tcagtcttta	gggggtgggc	tggtatgccg	ggtaaaaagt	ctttttgtct	taaaagaaaa	15840
aggaacatagg	tcaaaaaatc	gtccgtgacc	tatcagttat	taatttttaa	ggatgttgcc	15900
actgggcaat	atgtatgtgc	cagttcttttc	cataataaaa	ggctttagat	taactcaactg	15960
agggttactctg	acaatgtcga	ggttatgaac	aaagttagga	gaatgaaatt	tatgtgctct	16020
tagcaaaaac	atgtatgtgc	atttcaatcc	cacgtactta	taaaagaaagt	tggtagaattt	16080
cacaagctat	ttttggaata	tttttagaat	attttaagaa	tttcacaagc	tattccctca	16140
aatctcgagg	agctcagtaa	caccatcgat	catgatgtag	agtggtggtt	tgaactttaa	16200
agttatagtt	gttttatatg	ttgctataat	aaagaagtgt	ctcgacttcg	tcacgcgttt	16260
gttcattctg	tactgccact	tatctgctca	gttccctctc	aaaatagatt	aaagaactct	16320
ccttaagtaa	acatgtgtcg	tattctggtt	tggtatgtac	ttaaaagagt	atattttaga	16380
aataatagtg	aataatattt	gccctatttt	tctcatttta	actgcactct	atcctcaaaa	16440
tataatgacc	atttaggata	gagttttttt	tttttttttt	taaaccttta	taacctttaa	16500
gggttatttt	aaaaataatc	atggactacc	attttgcctt	cattagcttc	agcattggtg	16560
gactctctca	ataatatgct	tagattaagc	aaggaaaaaga	tgcaaaaacca	cttcgggggt	16620
aatcagtgaa	tattttttcc	cttcgttgca	taccagatac	ccccgggtgt	gcacgactat	16680
ttttattctg	ctaattttat	acaagtgtta	aacagaaaca	ggaattattc	caacacagtt	16740
tgcaacactgt	tgcttatttt	caaattacag	tttaatgtct	aggtgccagc	ccttgatata	16800
gctatttttg	taagaacatc	ctctcggaat	ttgggttagt	taaatctaaa	cttatttaag	16860
gattaagtag	gataacgtgc	attgattttg	taaaagaatc	aagtataaat	tacttagctg	16920
attcctgagg	gtggatattg	ttctagctga	actcatcttg	atcggtagga	ttttttaaat	16980
ccatttttgt	aaaaactttg	ccaagaaatt	ttaaagccct	tcacttcaga	aagaaaaaag	17040
ttgtgggggc	tgagcactta	attttcttga	gcaggaaagg	gtttcttcca	aacttcacca	17100
ttcggagact	gggtgtttct	tacagattcc	tccttcatct	ctgttgagta	gccgggatcc	17160
tatcaaaagc	caaaaaaatg	agtcctgtta	acaaccacct	ggaacaaaaa	cagattttat	17220
gcatttatgc	tgctccaaaga	aatgctttta	cgcttaagcc	agaggcaatt	aattaatttt	17280
tttttttttt	acatgggaat	actgtcogtt	gccacaggct	cagtgcaagt	gcgcgaatct	17340
ggctcactgc	aacctccacc	tcccagggtc	aagtgaattc	ctgtcctcag	ctctccatgt	17400
agctgggact	acaggccact	gccaccatgc	cgggctaatt	ttttgtattt	tttttagaga	17460
cagggtttca	ccatgttgcc	caggctggtc	tcaaacacot	gaacctcaat	gatccactctg	17520
cctcagcctc	ccaaagtgtt	gggattacag	gcgtaagcca	ccatgccccg	ccctgaatta	17580
atatttttaa	aataagtttg	gagactgttg	gaaataatag	ggcagaggaa	catattttac	17640
tggtcacttg	ccagagttag	ttaactcact	aaactcttgg	ataaatgatt	gacctcttgt	17700
gttgaaaaat	agccatgatc	tcttgaacat	gatcagaata	aatgccccag	ccacacaaat	17760
ggatgccaaa	cttttttagt	cactaacttg	ctagatgggt	ccaggttttt	ttgcacaagg	17820
agtgcaaatg	ttaagactct	cactagttag	gaaaggctag	tattacagaa	gctctgtcag	17880
agggcaattg	acctccaagc	cctggccctc	agggctgagg	attttgatga	agcacaactg	17940
aaagaacctg	gtttagtggg	tattgcaaac	aaacaggagt	caaagctgtg	tgctccacag	18000
tctagtctac	gagacaggcg	tgccagtggc	tgccagcatc	ttctctcaca	ggggccctca	18060
ggcacagctt	accttggggg	gcattgttaga	agcccgctgg	atcattcaagg	gatacttgaa	18120
atgtctcatg	aggttggtcaa	catactcaca	caccctagga	ggagggaatc	agatcggggc	18180
aatgatgcct	gaagtcagat	tattccagtg	gtgtcaactt	gagcagcaag	gagcgagtac	18240
cactcaacgc	acagtgtttg	ccaaggctta	gctgtgttac	catgcgtttc	taggcgaagt	18300
ccataaaact	tgctgcctcag	gtccttttct	tctaaaaaat	agcaattgtg	gggtggggact	18360
ttgatgacat	gaacacacga	agtcctctct	agaggttttg	tggtgccctt	taaaagggat	18420
caattagacat	tcgttaacat	tccagaaltt	tttgggttcc	agtcagataa	agtcagatga	18480
atagattaaa	atcaccacat	tttgtgatct	atttttcaag	aagcgtttgt	attttttcat	18540
atggctgcag	cagctgccag	gggcttgggg	tttttttggc	aggtagggtt	ggggagg	18596

<212> DNA
<213> Homo sapiens

<400> 12

accgggcaag	cgggaaaccag	gtggccacc	gggtgcggt	tcattttct	ttggaatttc	60
tgctttacag	acagaacaat	ggcagccga	gtacttataa	ttggcagtg	aggaagaggaa	120
catacagctg	cctggaaact	tgcaagctc	catcatgtca	aacaagtgtt	gggtgcccca	180
ggaaacgcag	gcactgcctg	ctctgaaaag	atctcaaata	ccgccatctc	aatcagtgac	240
cacactgcgc	ttgctcaatt	ctgcacaag	aagaaaaatt	aatttgtagt	tgtttggacca	300
gaagcacctc	tggtctgctg	gattgttggg	aacctgaggt	ctgcagagct	gcaatgcctt	360
ggcccaacag	cagaagcgcg	tcagtttaga	tcacgcaaaa	ggtttggcaa	agagtttatg	420
gacagactag	gaatcccaac	cgcacaatg	aaggctttca	ccaaaacctg	agaagcctgc	480
agcttcattt	tgagtcgaga	cttccctgct	ttggtttgta	aggccagtg	tcctgcagct	540
ggaaaagggg	tgattgttgc	aaagagcaaa	gaagaggcct	gcaaacgtgt	acaagagatc	600
atgcagagga	aagcctttgg	ggcagctgga	gaacaatttg	tcattgaaga	actctcttgac	660
ggagaagagg	tgtcgtgtct	gtgtttcact	gatggcaaga	ctgtggcccc	catgccccca	720
gcacaggacc	ataagcgatt	actggaggga	gatggtggcc	ctaacacagg	gggaatggga	780
gcctattgtc	cagccctcct	gggtttcta	gatctattac	taaaaaatt	agatactgtt	840
cttcagagga	cagtggaaag	catgcagcaa	gagggttact	catatacagg	tattctctat	900
gctggaataa	tgctgaccaa	gaatggccca	aaagtcttag	agtttaattg	ccgttttggc	960
gattccactg	gccaaagta	cctcccaatt	ctctaaaagt	atctttatga	agtgtattcg	1020
tcacacctag	ttaggactct	ctgcacatct	ctgcctgttt	ggctagaaaa	ccacacgcgc	1080
ctaaactgtt	tcattgcagc	taaaagttat	cctggagact	acacaaaggg	tgtagataga	1140
acagggtttc	ctgaggctca	agctctagga	ctggaggtgt	tcctatcgag	cactgcctct	1200
aaaaaatgca	aagttagta	tcattggggg	agagttcttg	cagtcacagg	catccgggaa	1260
aatctcatat	cagcccttga	ggaagccaa	aaaggactag	ctgtataaaa	gtttgaggga	1320
gcaatttata	ggaagacgt	cggctttcgt	gccatagett	tcctccagct	gcccaggagt	1380
ttgactatca	aggaactctg	agtagatctc	atagctggaa	atatgcttgt	caagaaaaatt	1440
cagcctctct	caaaaagccac	ttccagatca	ggctgtaaag	ttgatctctg	aggttttgc	1500
ggctcttttt	atttataaag	agctgggttc	aaagatcccc	ttctggccct	tggaaacagt	1560
ggcgttggaa	ctaaactaaa	gattgcccag	ctatgcaata	aacatgatac	catttggtcaa	1620
gatttggtag	caatgtgtgt	taatgatatt	ctggcacaag	gagcagagcc	ccctctcttc	1680
cttgattact	ttctctgtgt	aaaacttgac	ctcagtgtaa	ctgaagctgt	tgttgctgga	1740
attgtctaa	cttctggaaa	agctggatgt	gctctccttg	gaggtgaaac	agcagaaaatg	1800
cctgacatgt	atccccctcg	agagtatgac	ctagctgggt	ttgcctgttg	tgccatggag	1860
cgagatcaga	aactccctca	cctggaaaag	atcactgagg	gtgatgttgt	tgttggaata	1920
gcttcactct	gtctctctag	caatggattt	agcctttgtg	ggaaaaatct	tgcaaaaatc	1980
tcctccagct	actcctctcc	agcacctgat	gggtttgggt	accagacttt	aggggactta	2040
cttctccagc	ctacagcaat	ctacagccat	tcactgttac	ctgtctctac	ttcaggacat	2100
ctcaaaagct	ttggccattc	tactggtgga	ggattactag	agaacatccc	cagagctctc	2160
gtcgaaaact	ttggggtaga	tttagatgcc	cagacactga	ggatccccag	ggttttctca	2220
tggttgcaag	aggaagagca	cctctctgag	gaagagatgg	ccagaaactt	taactgtggg	2280
cttggcgctg	tcctttgtgt	atcaaaagga	cagacagagc	agattctgag	ggatatccag	2340
cagtcacaag	aagaagcctg	gggtgattgg	agtggtggtt	cacgagctga	aggttccccca	2400
cggtgtgaag	tcaagaactg	gattgaaagc	atgcaaaaat	gttggctcgt	cttgaaagaat	2460
ggctccctga	caaatcattt	ctcttttgaa	aaaaaaaagg	ccagagtggc	tgtcttaata	2520
tcctggaaag	gatcgaaact	gcaagcactt	atagacagta	ctcgggaagc	aaatagctct	2580
gcacaaaatt	atatttttat	ctccacaaca	gccgcagtag	ctgggttaga	taagtcggaa	2640
agagctggta	ttccactagt	agtaattaat	cataaactgt	ataaaaaatc	tgtagaattt	2700
gacagctgca	ttgacctagt	ccttgaagag	ttctccatag	acatagctct	ctctgcagga	2760
tcctatgaaa	ttctttttct	cccccttctc	caaaagtggg	atggaaaaat	gctcaaatc	2820
caccactcct	tgtctcctct	tttttaaggt	ttctaatgcc	atgagcaagc	ctctgaaacc	2880
ggagtcacag	ttactgggtg	cactgtacac	tttgtagctg	aagatgtgga	tgtctggacag	2940
attatttttg	aagaagctgt	tcctcgtaag	aggggttgata	ctgtcgcAAC	tcctttctgaa	3000
agagtataat	tagcagaaaa	taaaatattt	cctgcagccc	ttcagctgtt	ggccagtgta	3060
actgtacagc	ttgtgagaaa	tggcagaagt	tgtttgggtta	aagaggaatt	aaagctttta	3120
atttcagaaat	tgggccaggt	tagaagaat	tatttgtctg	tatttgtgtg	gtttttttatc	3180

CG27266.6.55560

tgtatgccaga	gaagatgagc	tccgagaagg	tggacgtgct	gcgggacactg	ggggctgaga	720
ttgtgaggac	gccacacaat	gccaggttcg	actccccgga	gtcacacagt	ggggtggcct	780
ggcggctgaa	gaacgaaatc	cccaattctc	acatcctaga	ccagtagccg	aacgccagga	840
accgccctggt	tactacagac	accaccgtgt	atgagatcct	gcagcagtg	gatgggaagc	900
tggacatgtg	gtgtggcttca	gtggggacgg	gcggccacat	cacgggcatt	gccaggaagc	960
tgaaggagaa	gtgtccctgga	tgcaggatca	ttgggggtgga	tcccgaaagg	tccatcctcg	1020
cagagccgga	ggagctggaac	cagacggagc	agacaacctc	gcaggttgaa	gggactcggt	1080
acgacttcat	ccccacgggt	ctggacagga	cggtgggtgga	caagtgggtc	aagagcaacg	1140
atgaggaggc	gttcaccttt	gcccgcatgc	tgatcgcgca	agaggggctg	ctgtgcgggt	1200
gcagtgcctg	cagcaggggt	gcggtggcgg	tgaaggctgc	gcaggagctg	gcaggagggc	1260
agcgctcggt	ggctcattctg	cccgcactag	tgcggaaacta	catgaccag	ttcctgagcg	1320
acaggctgat	gcctgagaag	ggctttctga	aggaggagga	cctcacggag	aagaagccct	1380
gggtgtggca	cctccgtgtt	caggagctgg	gcctgtcagc	cccgcgtacc	gtgctcccga	1440
ccatcacctg	tgggcacacc	atcgagatcc	tccgggagaa	ggggttcgac	caggggccgg	1500
tggtggatga	ggcgggggta	atcctgggaa	tgtgtacgct	tgggaacatg	ctctcgtccc	1560
tgcttcggcg	gaagggtcgag	ccgtcagacc	aagttggcaa	agtcatctac	aagcagttca	1620
aacagatccg	cctcacggac	acgctgggca	ggctctcgca	catcctggag	atggaccact	1680
tcgcctcggt	gggtgcacgag	cagatccagt	accacagcac	cggaaggtcc	agtcagcggc	1740
agatgggtgt	cgggggtggt	accgccattg	acttgctgaa	cttcgtggcc	gcccaggagc	1800
gggaccagaa	gtgaagtcgc	gagcgctggg	cggtgcggag	cgggcccgcc	acccttgccc	1860
acttctcctt	cgctttctctg	agccctaaac	acacgcgtga	ttggtaactg	cctggcctgg	1920
cacgcgttatc	ctctgcagagc	gcacagagca	tcctgtctccc	ctcgttaaca	catggcttcc	1980
taaatggccc	tgtttacggc	ctatgagatg	aaatatgtga	ttttctctaa	tgtaaactcc	2040
tcttaggatg	tttcacaaag	gaaatatgta	gagagaagtc	ggccaggtag	gatgaacaca	2100
ggcaatgact	gcgcagagcg	gattaaagcg	aaaagagaga	agagtccagg	aaggggccgg	2160
gagaagcctg	gggtggctcag	catcctccac	ggggtcgcg	ctgtcctcgg	gctgagctgg	2220
cgggagcagt	ttgcgtgttg	gggtttttta	attgagatga	aattcaaaat	acctaaaaat	2280
caatcacctg	aaagtgaaca	atcagcgcca	tttagtacat	ccagaaggtt	gtgtgagcac	2340
cacctctgtc	acgtttctgga	acattctgtc	atcaccccg	gaagacaatc	tttccctctc	2400
cgctctctct	ctccccctgc	aactgctgat	gcactttgtg	tctctgtgtg	ctaaaaatag	2460
ttttccctgt	tctggacatt	tcatataaat	ggaatcacac			2500

<210> 15

<211> 2068

<212> DNA

<213> Homo sapiens

<400> 15

cggcagccct	cctacctgcg	cacgtgggtc	cgctgctgct	gcctcccctg	cgccctgaac	60
ccagtgccctg	cagccatggc	tcccggccag	ctcgcttat	ttagtgtctc	tgacaaaaac	120
ggccttgttg	aatttgtcaag	aaacctgacc	gctcttggtt	tgaatctggt	cgcttccgga	180
gggactgcga	aagctctcag	ggatgtcgtg	ctggcagtc	gagatgtctc	tgagtgcag	240
ggatttctct	aaatgttggg	gggacgtgtg	aaaactttgc	atcctgcagt	ccatctgga	300
atcctagctc	gtaatatccc	agaagataat	gctgacatgg	ccagactgtg	tttcaactct	360
ataaaggttg	ttgctctcaa	tctctatccc	tttgtaaaga	cagtggtctc	tccaggtgta	420
actgtttgag	aggctgtgga	gcaaaattgac	attggtggag	taacctta	gagagctgca	480
gccccaaaa	acgctcgagt	gacagtgggt	tgtgaaccag	aggactagct	gggtgggtgc	540
accgagatgc	agagctccga	gagtaaggac	acctccttgg	agactagacg	ccagtagacc	600
tgaaggcat	tcaactatac	ggcacaatat	gatgaagcaa	tttcagatata	tttcaggaaa	660
cagtagacga	aaggcgtatc	tcaagatgcc	ttgagatatg	gaatgaaccc	caatcagacc	720
ctctcccgag	tgtacacact	gcagcccaag	cttcccatca	cagttctaaa	tggagccctt	780
ggatttataa	acttgttcga	tgttttgaa	gcctggcgac	tggtgaagg	actcaaggag	840
gcttttaggt	ttccagccgc	tgctcttttc	aaacatgtca	gcccagcagg	gtgctgtggt	900
ggaaatccac	tcagtgtaaga	tgaggccaaa	gtctgcatgg	tttatgatct	ctataaaaac	960
ctcacaccga	tctcagccgc	atatgcaaga	gcaagagggg	ctgataggat	gtcttcatatt	1020
gggtattttg	ttgcatttgc	cgatgtttgt	gatgtaccaa	ctgcaaaaat	ttcttccaga	1080
gaagtatctg	atgtgtataat	tgccccagga	tatgaagaag	aagccttgac	aatactttcc	1140

aaaaagaaaa	atggaaacta	tgtgtcctt	cagatggacc	aatcttacia	accagatgaa	1200
aatgaagttc	gaactctctt	tgggtctcat	ttaagccaga	agagaaataa	tgggtgcgtc	1260
gacaagttat	tatttagcaa	tgtgtgtacc	aaaaataaag	atttgcacga	gtctgccctc	1320
cgagacctca	tcgtagccac	cattgtgtgc	aagtacactc	agtcataact	tgtgtgcctc	1380
gccagaacg	ggcaggttat	cggcatttga	gcggacagc	agtcctgtat	acactgcact	1440
cgccctgcag	gagataaggc	aaactattgg	tggcttagac	accatccaca	agtgtcttcg	1500
atgaagttta	aaacaggagt	gaagagagca	gaaatctcca	atgccatcga	tcaatatgtg	1560
actggaaacca	tggcgaggga	tgaagatttg	ataaagtggg	aggcactgtt	tgagggaagtc	1620
cctgagttac	tcactgaggg	agagaagaag	gaatgggttg	agaaactgac	tgaagtttct	1680
atcagctctg	atgcctctct	ccctttccga	gataacgtag	acagagctgt	aaggagtggg	1740
gtggcgctaca	tgtgggtctc	ctccggttct	gctgctgaca	aagtgtgtgat	tgaggcctgc	1800
gacgaactgc	gaatcactct	cgctcatac	gaacctcggc	tcttccacca	ctgattttac	1860
cacacactgt	tttttggctt	gcttatgtgt	aggtgaacag	tcacgcctga	aactttgagg	1920
ataacttttt	aaaaataata	aacagtatct	cttaaaacaa	tgttttgatc	tcataaaca	1980
ttgtaaaaaa	tttcaatcac	gctttttaac	tttcttacca	caaaaaaatg	ataagtgggt	2040
gaagtgatgg	ttatgttaat	tagcgtgc				2068

<210> 16

<211> 857

<212> DNA

<213> Homo sapiens

<400> 16

gcgtgggcgt	gagatggcgg	cggcagcggg	gagcagcgcc	aagcggagcc	tgcggggaga	60
gctgaagcag	cgctgcgggg	cgatgagtgc	cgaggagcgg	ctacgcgcgt	cccgcgtact	120
gagccagaag	gtgattgcc	acagttagta	tcaaaagtcc	aaaaagaatt	ccatctttct	180
gagcatgcaa	gatgaattg	agacagaaga	gatcatcaag	gacatttttc	aacgagccaa	240
aatctgtctc	atccctcgtt	accggttcca	gagcaatcac	atggatatgg	tgagaataga	300
atcaccagag	gaaatttctt	tacttcccaa	aaatcctcgt	aatatccctc	agcctgtgtg	360
gggtgatggt	cgggaggagg	ccttgtccaa	agggggactt	gatctcactt	tcattgcagg	420
ccttgggttt	gacaacaatg	gcaaccgact	ggggaggggc	aagggtcact	atgatgccta	480
tctgaagcgc	tgtttgcagg	atcaggaagt	gaagccctac	accctggcgt	tggctttcaa	540
agaacagatt	tgcctccagg	tcaccagtga	tgaanaacac	atgaaggtag	atgaagtcct	600
ttacgaagac	tcgtcaacag	cttaaatctg	gattactaca	gccaaataat	cagtgtttta	660
tatgagagta	aagcaaatga	tgtgtatttt	tccttctgca	aaaattagtt	gaaattgttt	720
atlaattgtg	atcacagagt	catttttaaaa	ttgtaattat	gaaataacct	atataaaacc	780
attcttataaa	accaatagaa	gtgtgaatag	tagaataatta	attaaaaatg	aggctatcag	840
cctgtgattt	tcagctt					857

<210> 17

<211> 3762

<212> DNA

<213> Homo sapiens

<400> 17

cccgcgagcg	tcacatccat	tgtccggccg	actgtccagc	gaaaggggct	ccaggccggg	60
cgcacgtcga	cccggggggg	cgaggccagg	agagggggcca	agagcgcggc	tgacccttgc	120
ggggcggggc	agggggcggg	ggccgcggcc	atgcagtctt	gtggccgggt	gtggggggctg	180
cgcttggggc	ggcgggttcg	ggcgggccgc	cgcttggctg	ggggatcggt	gcccgtctgtg	240
ggcgccggga	ggcggggacg	cagcagtggt	ggcggggaca	ggcgccggcc	gtggggcctcg	300
cgcttctctg	agcgccttct	gcccgacacg	gacagcttgc	ctcggaggga	catcgccctc	360
ggggacaaaag	accagagaga	gatgtctgac	accttggggc	tggcgagcat	tggatgaattg	420
atcgagaaga	cggttcctgc	caacatccgt	ttgaaaaagc	ccttgaataa	ggaagaccct	480
gttctgtgaa	atgaaatcct	tgcaactctg	catgccattt	caagcaaaaa	ctgagctctg	540
agatcgtata	ttggcatggg	ctattataac	tgtctcagtc	cacagacgat	tttccggaac	600
tactctgaga	actcaggatg	gatcaccag	tatactccat	accagcctga	gggtgtctcag	660
ggggagcttg	agagtttact	caactaccag	accatggtgt	gtgacatcac	aggcctggac	720

atggccaatg	catccctgct	ggatgagggg	actgcagccg	cagaggcact	gcagctgtgc	780
tacagacaca	acaaggaggag	gaaatttctc	gttgatcccc	gttgccaccc	acagacaata	840
gcctgtgttc	agactcgagc	caaatatact	ggagtctctca	ctgagctgaa	gttaccctgt	900
gaaatggact	tcagtggaag	agatgtcagt	ggagtgttgt	tccagtaacc	agacacggag	960
gggaagggtg	aagactttac	ggaactctgt	gagagagctc	atcagagtgg	gagcctggcc	1020
tgctgtgcta	ctgacctttt	agcttttgtc	atcttgaggc	cacctggaga	atttggggta	1080
gacactgcgc	tgggcgctc	ccagagattt	ggagtgccac	tggtgtatgg	gggaccccat	1140
gcagcatttt	ttgtgtccg	agaaagcttg	gtgagaatga	tgctgggaag	aatgggtggg	1200
gtaaacaag	atgccactgg	gaaagaagt	tatcgtcttg	ctcttcaaac	caggggagcaa	1260
cacattcgga	gagacaaggc	taccagcaac	atctgtacag	ctcaggccct	cttggcgaa	1320
atgggtcgca	tgttttcgat	ctaccattgt	tccactgggc	tgagcatat	tgctaggagg	1380
gtacataatg	ccactttgat	tttgtcagaa	ggcttcaagc	gagcagcgca	tcaactccag	1440
catgacctgt	ctcttgatgc	cttgaagatt	catgtgggct	gtcagtgtaa	ggaggtcttg	1500
ggcaggcgcg	ctcagcgcca	gatcaatttt	cggttttttg	aggatggcac	acttggatt	1560
ctctcttgat	aaacagtcaa	tgaaaaagat	ctggacgatt	tgttgtggat	ctttggttgt	1620
gagtcactcg	cagaaactggt	tgctgaaagc	atgggagagg	agtgcagagg	tattccaggg	1680
ctctgtttca	agaggaccag	cccgttcttc	accatcaag	tgttcaacag	ctaccactct	1740
gaaacaaca	ttgtccggta	catgaagaaa	ctggaaaaata	aagacatttc	ctttgtttac	1800
agcatgatc	cactgggata	ctgcaccatg	aaactgaaca	gttctgtcga	actcgcaact	1860
atcacatgga	aagaatttgc	aaacatccac	ccctttgtgc	ctctggatca	agctcaagg	1920
tatcacagc	ttttccgaga	gcttgagaag	gattttgtgt	actcacagg	ttgtaccag	1980
gtctgtttcc	agccaaacag	cggagcccg	ggagaatatg	ctggactggc	caactaccga	2040
gcctacttaa	accagaaagg	agagggggc	agaaocggtt	gcctattcc	gaaatcagca	2100
catggggaca	accagacaag	tgccccatg	gcaggcatga	agattccagc	tggtggagtg	2160
gataaattcg	ggaatatcga	tgcatgttcc	ctcaaggcca	tggtggatga	gcacaaggag	2220
aaactagcag	ctatcatgat	tacataccca	tccaccaatg	gggtgtttga	agagaacatc	2280
agtgcgtgt	gtgacctcat	ccatcaacat	ggaggacagg	tctacctaga	cggggcaaat	2340
atgaatgctc	aggttgccct	ctgtgcacct	ggagacttcg	ggtctgatgt	ctcgacccta	2400
aatcttcaca	agacaccttg	catcccccac	ggaggagggt	gtctcggcat	ggggcccatc	2460
ggagtgaaga	aacatctcgc	cccgtttttg	cccatcatc	ccgtcatttc	actaaagcgg	2520
aatgaggatg	cctgtctctg	gggaaccgct	agtgcggccc	catggggctc	cagttccatc	2580
ttgcccattt	ctggggctta	tatcaagatg	atgggaggca	agggtcttaa	acaaggccag	2640
gaaactcgga	tattaaatgc	caactacatg	gccaaagcat	tagaaacaca	ctacagaatt	2700
cttttccagg	gtgcaagagg	ttatgtgggt	catgaattta	ttttggacac	gagacccttc	2760
aaaaagcttg	caaataatga	ggctgtggat	gtggccaaga	gactccagga	ttatggattt	2820
cacgccccta	ccatgtctcg	gcctgtggca	gggaccccca	tggtggagcc	cactgagtcg	2880
gaggacaagg	cagagcttga	cagattctgt	gatgccatga	tcagcatctg	gcaggaaatt	2940
gctgacattg	aggaggggcg	catcgacccc	agggtcaatc	cgtcgaagat	gtctccacac	3000
tcctcgacct	gcgtttacatc	ttcccactgg	gaccggccct	attccagaga	gggtggcaga	3060
ttcccactcc	ccttcatgaa	accagagaac	aaattctggc	caacgattgc	ccggattgat	3120
gacatatatg	gagatcagca	cctgggtttg	acctgcccac	ccatggaga	ttatgagctc	3180
ccattttctg	aacaaaagag	ggcgctcttc	tagtctcttc	tccttaagtc	taaaagagct	3240
atttgatgct	gttccccaga	gcatttgata	agcaagaaag	atttcatctc	ccaccaccgc	3300
ctcaagtagg	agttttatata	actgtgtata	tctctgtaat	ctctgtcaga	gttaaatgtaa	3360
atacagtagc	tgaggaggat	cgaagctgat	ggttgggaaga	cggattttgt	ttgttatctt	3420
gcttccacat	gtgccagatt	cctggatttg	gagccatttt	gtgtttttgc	tagaaaagtt	3480
taggaacatt	aactcttaata	gtggcaagtg	tcagatgtgc	atagaggctc	tcctggagac	3540
ttaatagaca	ttttttttgt	ccaaaagagt	ccatgtggac	tgctgccatc	gtgggaaatt	3600
ccaggggaaa	tgttttacatt	ttgtataacc	tgaagaactc	tttttctctt	aatatgccta	3660
atctgataac	acatttctga	gtgttttctc	gtgttttctg	gtgaggtttt	tttttttttt	3720
aatctgcatt	tattagattt	ctaataaaag	cattttgatc	gg		3762

<210> 18
 <211> 1192
 <212> DNA
 <213> Homo sapiens


```

ggagattctg cctttgccag gtgactgtct ggggagcggg tctgtcccca aggggcctga 2640
gcagtccttg gctgtctaag gtcttggaac ttgcctgcct ttccatccat ggccagcagc 2700
acctgcccta cctgccccac ttgtccttag cctggacctc tgacagcagc atctctacct 2760
tctccccagc tccaggagac acaggctcag gcagggcctc catggggccc aggggaacac 2820
tggggacttg gctctctctc aggggtacatg gtgctgggag aggcagccca ggaagtctca 2880
tctggggagc aggcagccag catctgggcc ttggcctgga gcacaaagac cctggctttc 2940
attttctctc aggtgaaagg aaattaaggc aacaaaaagaa gcccggtccc tggtcaccta 3000
ggaagcctca gattccttcc catggaggga gggagtgggt tgcagggtgc caagtctctc 3060
taacttggct cacactgcag atgaaaattc agaattttat actttcccta ccctctagag 3120
aaataagatc ttttttgtca gtttgtttgt atgaaactaa agctttattt gttaataagt 3180
cctgctaaaa caatgaataa aaactcaagg agcaactaaa aaaaaaaaa 3228

```

```

<210> 21
<211> 344
<212> PRT
<213> Homo sapiens

```

```

<400> 21
Met Ser Ala Leu Ala Ala Arg Leu Leu Gln Pro Ala His Ser Cys Ser
  1             5             10             15

Leu Arg Leu Arg Pro Phe His Leu Ala Val Arg Asn Glu Ala Val
  20             25             30

Val Ile Ser Gly Arg Lys Leu Ala Gln Gln Ile Lys Gln Glu Val Arg
  35             40             45

Gln Glu Val Glu Glu Trp Val Ala Ser Gly Asn Lys Arg Pro His Leu
  50             55             60

Ser Val Ile Leu Val Gly Glu Asn Pro Ala Ser His Ser Tyr Val Leu
  65             70             75             80

Asn Lys Thr Arg Ala Ala Ala Val Val Gly Ile Asn Ser Glu Thr Ile
  85             90             95

Met Lys Pro Ala Ser Ile Ser Glu Glu Glu Leu Leu Asn Leu Ile Asn
 100             105             110

Lys Leu Asn Asn Asp Asp Asn Val Asp Gly Leu Leu Val Gln Leu Pro
 115             120             125

Leu Pro Glu His Ile Asp Glu Arg Arg Ile Cys Asn Ala Val Ser Pro
 130             135             140

Asp Lys Asp Val Asp Gly Phe His Val Ile Asn Val Gly Arg Met Cys
 145             150             155             160

Leu Asp Gln Tyr Ser Met Leu Pro Ala Thr Pro Trp Gly Val Trp Glu
 165             170             175

Ile Ile Lys Arg Thr Gly Ile Pro Thr Leu Gly Lys Asn Val Val Val
 180             185             190

Ala Gly Arg Ser Lys Asn Val Gly Met Pro Ile Ala Met Leu Leu His
 195             200             205

```

Thr Asp Gly Ala His Glu Arg Pro Gly Gly Asp Ala Thr Val Thr Ile
210 215 220

Ser His Arg Tyr Thr Pro Lys Glu Gln Leu Lys Lys His Thr Ile Leu
225 230 235 240

Ala Asp Ile Val Ile Ser Ala Ala Gly Ile Pro Asn Leu Ile Thr Ala
245 250 255

Asp Met Ile Lys Glu Gly Ala Ala Val Ile Asp Val Gly Ile Asn Arg
260 265 270

Val His Asp Pro Val Thr Ala Lys Pro Lys Leu Val Gly Asp Val Asp
275 280 285

Phe Glu Gly Val Arg Gln Lys Ala Gly Tyr Ile Thr Pro Val Pro Gly
290 295 300

Gly Val Gly Pro Met Thr Val Ala Met Leu Met Lys Asn Thr Ile Ile
305 310 315 320

Ala Ala Lys Lys Val Leu Arg Leu Glu Glu Arg Glu Val Leu Lys Ser
325 330 335

Lys Glu Leu Gly Val Ala Thr Asn
340

<210> 22

<211> 1283

<212> DNA

<213> Homo sapiens

<400> 22

tttcgcagcc	gctgccgcct	cgccgctgct	ccttcgtaag	gccacttccg	cacaccgaca	60
ccaacatgaa	cggaacagctc	aacggccttc	acgaggcggt	catcgaggag	ggcaccattcc	120
ttttcacctc	agagtcggct	gggggaaggcc	accagagataa	gattttgtgac	caaatcagtg	180
atgctgtcct	tgatgccccac	cttcagcagg	atcctgatgc	caaagttagct	tgtgaaactg	240
ttgctaaaac	tggaaatgac	cttccttgctg	gggaaattac	atccagagct	gctgttgact	300
accagaaagt	gggtcgtgaa	gctgttaaac	acattggata	tgatgattct	tccaaagggt	360
tgactacaa	gacttgttaac	gtgctggtag	ccttggagca	acagtcacca	gatattgtct	420
aagggtgtca	tcttgacaga	aatgaagaag	acattgggtgc	tggagaccag	ggcttaatgt	480
ttggctatgc	actgatgaa	actgaggagt	gtatgccttt	aaccattgtc	ttggcacaca	540
agctaaaatgc	caaactggca	gaactacgct	gtaatggcac	tttgctctgg	ttacgacctg	600
attctaaaac	tcaagttact	gtgcagtata	tcgaggatcg	aggtgctgtg	cttcccatca	660
gagtcacac	aattgttata	tctgttcacg	atgatgaaga	ggtttgtctt	gatgaaatga	720
gggatgcctt	aaaggagaaa	gtcatcaaa	cagttgtgct	tcgcaaatat	cttgatgagg	780
atacaattcta	ccacctacag	ccaagtggca	gatttgttat	tggtgggctt	cagggtgatg	840
ctgggttgac	tggacggaaa	atcattgtgg	acacttatgg	cggttggggg	gctcatggag	900
gaggtgcctt	ttcaggaaag	gattatacca	aggtcgaccg	ttcagctgct	tatgctgctc	960
gttgggtggc	aaaatccctt	gttaaaggag	gtctgtgccg	gagggttctt	gttcagggtct	1020
cttatgctat	tggagtttct	catccattat	ctatctccat	tttccattat	ggtagcctctc	1080
agaagagtga	gagagagcta	ttagagattg	tgaagaagaa	tttcgatctc	cgccctgggg	1140
tcattgtcag	ggatctggat	ctgaagaagc	caatttatca	gaggactgca	gcctatggcc	1200
actttgtag	ggacagcttc	ccatgggaag	tgcccaaaaa	gcttaaatat	tgaagtggtt	1260
agcctttttt	ccccagactt	gtt				1283

<210> 23
 <211> 3259
 <212> DNA
 <213> Homo sapiens

<400> 23

caaggttggt	ggaagtgcg	ttgtgcaggt	tcgtgcccg	ctggcgccg	gtggtttcac	60
tgttacatgc	cttgaagtga	tgaggaggtt	tctgttacta	tatgctacac	agcagggaca	120
gccaaagggc	atcgcagaag	aaatgtgtga	gcaagctgtg	gtacatggat	ttcttcgaca	180
ttctcactgt	attagtgaat	cgcataagta	tgacctaaaa	accgaacacg	ctctcttctg	240
tggtttgggt	tctaccacgg	gcaccggaga	cccaccggac	acagcccgca	agttttgttaa	300
ggaaatacac	aaccaaacac	tgccggttga	ttcttttgc	ccctgcgggt	atgggttact	360
gggtctcggt	gattcagaat	acacctactt	ttgcaatggg	gggaagataa	ttgataaacg	420
acttcaagag	cttggagccc	ggcatttcta	tgacactgga	catgcagatg	actgtgtagg	480
tttagaacct	gtggttgagc	cggtgattgc	tggaactctg	ccagccctca	gaaagcattt	540
taggtcaagc	agaggacaag	aggagataag	tggcgcactc	ccggtggcat	cacctgcatc	600
cttgaggaca	gaccttgtga	agtccagagc	gctacacatt	gaatctcaag	tcgagcttct	660
gagattcgat	gattcaggaa	gaaaggattc	tgaggttttg	aagcaaaatg	cagtgaacag	720
caaccaatcc	aatgttgtta	ttgaagactt	tgagctctca	cttacccggt	cggtaccctc	780
actctcacaa	gcctctctga	atattctctg	tttaccctca	gaatatattc	aggatcatct	840
gcaggagtct	cttggccagg	aggaaagcca	agatctctgt	acttcagcag	atccagtttt	900
tcaagtgtcca	atttcaaaag	cagttcaact	tactatcgat	gatgccataa	aaaccactct	960
gctgttagaa	ttggacaatt	caaatcacga	cttttctcat	cagccttgag	atgctctcag	1020
cgatgatctg	cctaacaggt	attctgaggt	acaaagccta	ctccaaagac	tcgagcttga	1080
agataaaaag	gagcactgcg	tccttttgaa	aataaaggca	gacacaaaag	agaaaaggagc	1140
taccttacc	cagcatatac	ctgcgggatg	ttctctccag	ttcattttta	ctggtgtctt	1200
tgaatccga	gcaattctta	aaaaggcatt	tttgcgagcc	cttgtggcat	ataccagtag	1260
cagttctgaa	aagcgcagcc	tacaggagct	gtgcagtaac	caaggggcag	ccgattatag	1320
ccgctttgtg	cgagatgcct	gtgcctgctt	gttggatctc	ctctctgctt	tcctctctgt	1380
ccagccacca	ctcagttctc	tgtctgaaca	tcttctctaa	cttcaaccca	gacctatctc	1440
gtgtgcaagc	tcaagtttat	ttcaccagg	aaagctccat	ttgtcttcca	acattgtgga	1500
attttctgtc	actgcccaca	cagaggttct	gcggaaggga	gtatgtacag	gctggctggc	1560
cttgtttggt	gcttcagttc	ttcagccaaa	catacatgca	tcctatgaag	acagcgggaa	1620
agccttggtc	cttaagatat	ccatctctcc	tcgaacaaca	aattctttcc	acttaccaga	1680
tgaccctcca	atccccatca	taattgtggg	tcagggaacc	ggcatagccc	cgttttattg	1740
gttctctaca	catagagaga	aactccaaga	acaacaccca	gatggaattt	ttggagcaat	1800
gtggttggtt	tttggttgca	ggcataagga	tagggattat	ctattcagaa	aagagctcag	1860
acatttctct	aagcatggga	tcttaactca	tctaaagggt	tccttctcaa	gagatgctcc	1920
tggtggggag	gaggaaagcc	cagcaaaagta	tgatacaagc	aaatctctca	ttcatggcca	1980
gcagggtggc	agaattctcc	tccaggagaa	cgcccatatt	tatgtgtgtg	gagatgcaaa	2040
gaattatggc	aagatgtgac	atgatgcctt	ttgtgcaata	ataagcaaa	aggttggagt	2100
tgaaaaacta	gaagcaatga	aaacccctgg	cacttttaaa	gaagaaaaac	gctaccttca	2160
ggatatattg	tcataaaacc	agaaatttaa	gaagaggat	taagcttttt	tgactgaag	2220
tactaaaaag	cagcttttact	agtgccaaac	cttttaattt	tcaaaagaaa	taatttcttc	2280
aacatttctt	gaaggacatg	gagtgagat	tggatcattt	acaataataa	caaaacttcc	2340
tgatttgatt	ttacgtatct	tctatctacg	cccttctctg	gctctgtagt	ctccccaaat	2400
tgccctgttg	ccttgagctt	ttctgagcta	aaggcagcct	tcagtcctca	tcagcgctcc	2460
ctttactctc	cagagaaact	cacagagact	ctgtccttcc	atgcgaaagg	ttcgtgaaat	2520
agggggagact	gactgtgtag	ctcattctgt	tgactctacg	tgcgaacatt	taaaaaagta	2580
tgaaaatgat	ttatttttat	atgatgtata	ccataaaga	ctctctat	tgctgtactt	2640
aaattcacaca	tgtagagcat	atctgttata	tgtttatgta	actatcaaat	gggttattgt	2700
tactaaaagt	atatttctga	taaaaaatat	tttaggataa	ttgctacag	agggatttat	2760
ttttatgatg	ctgggaataa	tgaaatgtat	tttaaaattt	cactctgggc	atatggattt	2820
atctatcacc	attacttttt	tttaagtcac	aatttcagaa	ttttgggaca	tttgcatcca	2880
atttacaggt	accagtgact	acatatttta	atagaaaagt	acaacctttt	tattttcaat	2940
cttttatttt	ctgctctctg	gcacattttt	gagttttccc	acattatttt	ttctccatgt	3000
accactcaag	cagtggtgctg	gacctaaaa	actgacttta	gttagtatcc	ttggattttt	3060

CG577666.103333

agattcccca	gtgtctaatt	cctgtttata	atttgacaaa	acaaaacaaa	atgttatgat	3120
aatctttctc	cactgttcta	atatataattg	tattttttatt	tgtatagcttg	ggatttaaaa	3180
cattctgtgt	gaaggctttt	gatccttttg	agaaaataaag	atctgaaga	aatggcataa	3240
tcttaaaaaa	aaaaaaaaa					3259

<210> 24
 <211> 1805
 <212> DNA
 <213> Homo sapiens

<400> 24						
aagagactga	actgtatctg	cctctatttc	caaaagactc	acgttcaact	ttcgctcaca	60
caaaagccggg	aaaattttat	tagtctcttt	tttaaaaaaa	gttaataataa	aattatagca	120
aaaaaaaaaaa	ggaacctgaa	ctttagtaac	acagctggaa	caatcgcgag	ggcgccggca	180
gcggcgggag	aagaggttta	atttagttga	ttttctgttg	ttgttggttg	ttcgctagtc	240
tcacggtgat	ggaagctgca	cattttttcg	aagggaccga	gaagctgctg	gaggtttggg	300
tctcccgcca	gcagcccgac	gcaaaccaag	gatctgggga	tcttcgact	atcccaagat	360
ctgagtggga	catacttttg	aaggatgtgc	aatgttcaat	cataagtgtg	acaaaaactg	420
acaagcagga	agcttatgta	ctcagtgaga	gtagcatgtt	tgctcccaag	agacgtttca	480
ttttgaagac	atgtgggtacc	accctcttgc	tgaagcact	ggttccctcg	ttgaagcttg	540
ctagggatta	cagtgggttt	gactcaattc	aaagcttctt	ttattctcgt	aagaatttca	600
tgaagccttc	tcaccaaggg	taccacaccc	ggaatttcca	ggaagaataa	gagtttctta	660
atgcaatttt	cccaaatgga	gcaggatatt	gtatgggacg	tatgaattct	gactgttggg	720
acttatatac	ctctggatttc	ccagagagtc	gggtaatcag	tcagccagat	caaaccttgg	780
aaattctgat	gagtgagctt	gaccagcag	ttatggacca	gttctacatg	aaagatgggt	840
ttactgcaaa	ggatgtcact	cgtgagagtg	gaattcgtga	cctgatacca	ggttctgtca	900
ttgatgccac	aatgttcaat	ccttgtgggt	attcgatgaa	tgggaatgaa	tcggatggaa	960
cttatgggac	tattcacata	actccagaac	cagaattttc	ttatgttagc	tttgaacaaa	1020
acttaagtca	gacctcctat	gatgacctga	tcaggaaaag	tgtagaagtc	ttcaagccag	1080
gaaaattttg	gaccaccttg	tttgttaatc	agagttctaa	atgtgcgaca	gtgcttgcct	1140
cgccccagaa	gattgaaggt	tttaagcgtc	ttgatggcca	gagtgctatg	ttcaatgatt	1200
acaattttgt	ttttaccagt	tttgctaaga	agcagcaaca	acagcagagt	tgatttaagaa	1260
aaatgaagaa	aaaaacgcaa	aagagaacac	atgtagaagg	tggtggatgc	tttctagatg	1320
tcgatgctgg	gggcagtgtc	ttccataacc	accactgtgt	agttgcagaa	agccctagat	1380
gtaatgatag	tgtaatcatt	ttgaattgta	tgcatattta	tatcaaggag	ttagatatct	1440
tgcatgaatg	ctctctctct	tgtttaggta	ttctctgcca	ctcttgcgtg	gaaattgaa	1500
tggtatgtag	aaaaaccttt	tactatatga	aactttacaa	cacttgtgaa	agcaactcaa	1560
tttgggttat	gcacagtgta	atatttctcc	aagtatcatc	caaaattccc	cacagacaag	1620
gctttcgtcc	tcattagggt	ttggcctcag	cctaaccctc	taggactgtt	ctattaaatt	1680
gctgcccagaa	ttttacatcc	agttactctc	actttctaga	acataattctt	tactaatggt	1740
attgaaacca	atttctactt	catactgatg	ttttggaaa	cagcaattaa	agtttttctt	1800
ccatg						1805

<210> 25
 <211> 254
 <212> PRT
 <213> Homo sapiens

<400> 25										
Gln Asp Ile	Leu Val	Phe Arg	Ser Lys	Thr Tyr	Gly Asn	Val Leu	Val			
1		5		10		15				
Leu Asp Gly	Val Ile	Gln Cys	Thr Glu	Arg Asp	Glu Phe	Ser Tyr	Gln			
	20		25		30					
Glu Met Ile	Ala Asn	Leu Pro	Leu Cys	Ser His	Pro Asn	Pro Arg	Lys			
	35		40		45					

Val Leu Ile Ile Gly Gly Gly Asp Gly Gly Val Leu Arg Glu Val Val
50 55 60

Lys His Pro Ser Val Glu Ser Val Val Gln Cys Glu Ile Asp Glu Asp
65 70 75 80

Val Ile Gln Val Ser Lys Lys Phe Leu Pro Gly Met Ala Ile Gly Tyr
85 90 95

Ser Ser Ser Lys Leu Thr Leu His Val Gly Asp Gly Phe Glu Phe Met
100 105 110

Lys Gln Asn Gln Asp Ala Phe Asp Val Ile Ile Thr Asp Ser Ser Asp
115 120 125

Pro Met Gly Pro Ala Glu Ser Leu Phe Lys Glu Ser Tyr Tyr Gln Leu
130 135 140

Met Lys Thr Ala Leu Lys Glu Asp Gly Val Leu Cys Cys Gln Gly Glu
145 150 155 160

Cys Gln Trp Leu His Leu Asp Leu Ile Lys Glu Met Arg Gln Phe Cys
165 170 175

Gln Ser Leu Phe Pro Val Val Ala Tyr Ala Tyr Cys Thr Ile Pro Thr
180 185 190

Tyr Pro Ser Gly Gln Ile Gly Phe Met Leu Cys Ser Lys Asn Pro Ser
195 200 205

Thr Asn Phe Gln Glu Pro Val Gln Pro Leu Thr Gln Gln Gln Val Ala
210 215 220

Gln Met Gln Leu Lys Tyr Tyr Asn Ser Asp Val His Arg Ala Ala Phe
225 230 235 240

Val Leu Pro Glu Phe Ala Arg Lys Ala Leu Asn Asp Val Ser
245 250

<210> 26

<211> 2211

<212> DNA

<213> Homo sapiens

<400> 26

ctgaggccca gcccccctcg cccgtttcca tcacgagtgc gcgccagcatg tctgacaaac 60
tgccctacaa agtcgcgcag atcggcctgg ctgcctgggg acgcaaggcc ctggacattg 120
ctgagaacga gatgccgggc ctgatgcgta tgcgggagcg gtactcggcc tccaagccac 180
tgaagggcgc ccgcatcgct ggctgcctgc acatgaccgt ggagacggcc gtccctcattg 240
agaccctcgt caccctgggt gctgaggtgc agtgggtccag ctgcaacatc ttctccacc 300
agaaccatgc ggcggctgcc attgccaagg ctggcattcc ggtgtatgcc tggaaagggcg 360
aaacggacga ggagtagctg tgggtgcattg agcagaccct gtacttcaag gacggggcccc 420
tcaacatgat tctggacgac gggggcgacc tcaccaacct catccacacc aagtaccgcg 480
agctctctgcc aggcattccga ggcattctctg aggagaccac gactggggct cacaacctct 540
acaagatgat ggccaatggg atcctcaagg tgctgtccat caatgtcaat gactccgtca 600

atagaattac	aaatagcact	tgataatfff	aaagtatggt	ttagaatfff	tcttaggagc	1440
aaaaaagta	caaagtaaat	cttgaacagg	ttcactaagc	accaccctgt	tgaagaagtat	1500
tattgaaatc	actgcgcac	agggaaaagta	attcagatgt	taatgccact	tgaagaagtat	1560
gttagggctag	caaagaggat	gagacatgaa	ctgtcataaa	ggactcagca	accagccagg	1620
gacagcctaaa	gcgctatgga	aaggggcttc	caagttcttt	tgaaactgac	ccttagtaac	1680
aaacacaatt	tatatataatga	cccagcaaaa	cacatcacat	cttactgtcg	aaattaaatg	1740
tgtgtcccat	cctagctatfff	tctgttccat	tcttttccat	tctatttcat	ttataaaaaca	1800
tgctagtgtga	gacttttcaa	atggatfff	atgaccact	actgggtgtg	gatccacagt	1860
ttgaaaaaa	ttgtacaag	acacttaagg	agaccatct	gttttaagtt	attcttataa	1920
gtaggtcagt	catatgagac	ctgatcaata	aatatccaat	accagagatt	ctgtctcag	1980
agttctctct	ttctgtgacc	cacttttcta	ccagtaaaag	acatagacca	atggggaggga	2040
ggggaggaga	gatggatatt	tcagccctct	ccatccctag	caacactctg	tccactagt	2100
gcctctgggc	cataaggctg	agcagagtga	gcttgtatta	gttggtagct	tttaaaaaat	2160
ataataaaaa	aaaagtagag	attctccaaa	ctctagcctg	gtttctcaga	ttgagaacta	2220
tgatatffff	ctctgataat	ttaatatcta	ctctctcaca	aaagctcaag	cctgaagata	2280
caagactatt	agaagaaaaa	tgactaccct	cagtgattta	gaaaagggt	catgcagctt	2340
tctaaacatt	attgaattgt	ttgagctgtt	ttgaattgt	aattcttttc	agctattaaa	2400
aagaagagca	atgagaaaaa	aaaaaaaaaa	aaaaaa			2436

<210> 28

<211> 1326

<212> DNA

<213> Homo sapiens

<400> 28

ttcttttctt	ctcttttctt	tctcggttct	agcattgcag	aaaaagagc	ctcctcaca	60
gggttctctg	cacacttcca	acatttctgc	acgcaggcga	tccattgtgg	ccaggatccg	120
gagcaatgga	cctccaggcc	tgtagtgcct	cccatctcac	tgctccaccac	gttccaagaca	180
ggggcgccgt	gccagcactc	gggttttgaa	tatagccgtt	ctggaatccc	cactaggaat	240
tgcccttgaaa	aagcagtggc	agcactggat	ggggctaaat	actgtttggc	ctttgtctca	300
gggtttagcag	ccactgtaac	tattaccctt	cttttaaaag	caggagacca	aattatttgt	360
atggatgatg	tgtatggagg	tacaacacag	tacttcaggc	aagtggcatc	tgaatttgga	420
ttaaagattt	ctttttgtga	tgtttccaaa	atcaaattac	tagaggcagc	aattacacaa	480
gaaaccaagc	ttgttttgga	cgaacccccc	acaaacccca	cccaagaggt	gattgacatt	540
gaaggctgtg	cacatattgt	ccataagcat	ggagacatta	ttttggtcgt	ggataaacatt	600
tttatgtcac	catatttcca	gcgccctttg	gctctgggag	ctgatatttc	tatgtattct	660
gcaacaaaat	acatgaatgg	ccacagtgtg	gttgtaattg	gcctgggtgc	tgtaatttgt	720
gaaagccttc	ataatagact	tcgtttcttg	caaaactctc	ttggagcagt	tccatctcct	780
attgattggt	acctctgcac	tcgaggtctg	aagactctac	atgtccgaat	ggaaaagcat	840
ttcaaaaaag	gaatggcagt	tgcccagttc	ctggaatcta	atcctttggg	agaaaagggt	900
atttatctct	ggctgccttc	tcactccacg	catgagttgg	tgaagcgtca	gtgtacaggt	960
gtgacaggga	tggtccacct	ttatattaa	ggcactcttc	agcatgtcga	gattttctct	1020
aagaacctaa	agctatttac	tctggccgag	ctctggggag	gattcgaaag	ccttgctgag	1080
cttcgcggca	tcagtactca	tgcatcagtt	cttaagaagt	acagagatgt	ccttgggaatt	1140
agtgacacac	tgattcgact	ttctgtgggc	ttagaggatg	aggaagacgt	actggaagat	1200
ctagatcaag	ctttgaagcg	agcacaccct	ccaagtggaa	ttcacagcta	gtattccaga	1260
gctgctatta	gaagctgctt	cctgtgaaga	tcaattcttc	tgagtaatta	atggaccaac	1320
aatgag						1326

<210> 29

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:PCR product

<400> 29
cccacggtcg gggtagctcg gcgggacgcg ccaggccgac tcccgcgga

49

<210> 30
<211> 3464
<212> DNA
<213> Homo sapiens

<400> 30
tttaattggac acataatttta attatatatt ttttcttaca gataccocagg tgttctctct 60
gatgtccagg agggagaaagg cattaagtagt aaatttgaag tatatgagaa gaattgattaa 120
tatgaagggt tttcttagtt taagtgtgtc cccctccctc tgaaaaaagt atgtattttt 180
acattagaaa aggttttttt ttgacttttag atctataatt atttctaagc aactagtttt 240
tattccccac tactctgtgt tctatcagat accattttat agacattctt gctataacta 300
agtgtctctc caagacccca actgagttccc cagcacctgc tacagttagc tgccattcca 360
caccatcac atgtggcact cttgccagtc cttgacattg tcgggctttt cacattgttg 420
taattatttat taaagatgaa gatccacata ccttccaact gaggcagtttc actagtggaa 480
ataccaaaag ctctctacgt gtatatccag aggtttgtag ataaattgtt ccaccttgtt 540
tgtaacagtg aaaaattgaa aacaacctgg aagctccagt tgggaaaagt gagtatgttt 600
ctgtcttaga ttggggaaac caaagcagat tgcaagactg aaatttcagt gaaagcagtg 660
tattgtctag gtcataccag aaatcatcaa ttgaggtacg gagaacctga actgagaagg 720
taagaaaagc aatttaaagt cagcgaagcag gttctcattg atacaagact ccatactgct 780
gagatcacagg gaaattgagg ggggaaagct ggagattga tccccccccc ctctctgttt 840
gtcagctccc tgtcctgtgt gtggcggaac catagtcagg ctgtctata gcaagtctca 900
gggtgttgca gtaagaagct gctggcatgc acgggaaacg tgaatgccaa acacttaagg 960
caattcagat ttaagtattg taagtctttt tttttttaga cagcgttttc ctctctgttc 1020
ccaggctagg atgcaatggt gtgacctcgg ctactcgcaa cctccgctt cccagattgc 1080
agcgattctc ctgcctcagg ctcccaagta gctaggacca ggtgcgcgcc accacgcccg 1140
gcataatttt gtatttttga ttttttagtag agatgggggt tccacattgtt ggtcaggcta 1200
gtctcgaaat cgtgaccgca agcgattcac ccacctcagc ctcccaaggt gctgggatta 1260
ccggcttgag ccaccacacc cggcacatct tcattctttt tatgtagtaa aaagtataag 1320
gccacacatg gtttatttga agtattttat aatttaaaaa aatcacagaag caggaaaaac 1380
aattataagt tcaagttagg gatgatggtt gcttgaacca aaggggttga ttagtagaa 1440
aattgtgatt taagatatat tttaaagtta taagtagcag gatattctga tggagtttga 1500
ctttgttttt gggcccgagg agtttcagat gcccttgaga aatgaatgaa gttagagaaa 1560
aataaaaaga aaaaacgcca ggcacagtgg ctcacacctg taatcccgag gctttgggag 1620
gctaaggcag gcagatcact tgagaccagc ttgggcaaca tggcaaaagc ccattctctac 1680
aaaaaacaca aaaattagct gggcatttgt gcgcacacct gtatcccat ctagtacaga 1740
agctgagatg gaagaattaa ttgagccac gagttcaagg ctgcagttag tcgtgattgt 1800
gccactgcac tcacgcccag gtgacagaag agaccttctc tcgaaaaaca atctgaaaaa 1860
aatggaacca tgccctcata attctagaaa gattatttca actgataaat ctatattcac 1920
cacaataatc aagggtagag gtaaaaatat acatttttag acaagcgaag actcaggggt 1980
tacctccatg tgcccctttt aggggaagctg ttggagaaaa tactccaaga aatatcagga 2040
gtacacaaac cagagaatga catgaatcca gcaaatagga tccaacacag gcaattattc 2100
agctatggag ctgcttttaa aaaggaaacg aatcggttag atcggttgaa ctgtgggaa 2160
tgcccatgct ctgtagtcac agctactcag gaggctcagc agcaggacga cttagaccca 2220
agagttccag accagcctgg ccaccttagt gagatccctt ctcttaaaaa taataactta 2280
tgccagatt tggggcattt ggaagaagat tcatgaaaga taagcaaaa gtaaaaaaaa 2340
aaaaaataaa aacaaagggg aagggttggt taggcaatca ttctagggca gaagaagtaa 2400
caggatagga agagcataat acactgtttt tctcaacaag gagcagtagt tacacagtca 2460
taatgatgtg actgcttagc ccttaaatat ggtaactact ctgggacaaat atgggaggaa 2520
aagtgaagat tgtgatgtgt taagagctaa tctctactat caatcacact aaatcacact 2580
ataatcatata ataattgaat gactaagtta tgtgaggaaa aaaaacagaa acattgctaa 2640
aagagttaaa agtctattgt ctggagaatt aggagggatg gggcagggga ctgtaggat 2700
gcattataaa ctgaaaagcc ttttttaaat tttatgtatt aatatatgca ttcacttgaa 2760
aaactaaaaa aaacataaaa tttggaaaaa ccatcgaagc taactacagc aaggaaaaac 2820
taagagaatg aaaagtattt gcctctggaa agaacaactg gcaggactgt tgttttcatt 2880

CCAGGCTCG GGGTAGCTCG GCGGGACGCG CCAGGCCGAC TCCCGCGGA

gtaagacttt	tggagccatt	taattgtact	taaccatttt	catctatttc	tttaataaga	2940
acaattccat	cttaataaag	agttacactt	gttaataagt	gctggccctc	tgtgtgtctt	3000
gtcacacccc	acacaaaatt	tcaagaatac	tttgatggca	atatctcttc	atgggcagct	3060
taaaaataga	gaagggaaaa	catagaatta	gccaaagagt	acacaaaaca	aagatcagtt	3120
gtttgttagg	aaacaattcaa	aatcaagtct	cacttttttc	agatttggct	atgggaacag	3180
actgttaagg	gataacttgg	ggcaaacatg	taaaataata	aacatattgt	ttaaatatct	3240
agggttagac	attttattgt	tctgttagat	taaaatttgg	tgtgacatac	ccgcttccct	3300
aaaggcaatg	ttttctgaaa	tgttgtacct	gctattctct	aatcagggat	gggtcccaga	3360
atctgccttt	taaacatctc	agataatctg	aagcctgctt	aagtttgtaa	ggcaactgct	3420
ttgcactcta	aggaagaaaa	aaacaagttt	taattcccg	ctct		3464

<210> 31
 <211> 1584
 <212> DNA
 <213> Homo sapiens

<400> 31						
cggggacgct	ctgaggaaca	aggttgaagc	tcagagcgct	ggctctccacc	ctgggtgccc	60
tgggtcggtg	ctggcagttg	gagccgtggc	tgtggatgag	agacatagac	gagagagtga	120
gatggccctg	tttgcctctt	acctcctgag	ccttctcttg	gctacagctg	ggactagtac	180
cgagaccag	agttctatgt	ccgttccctc	agcacaggag	cccttgggtca	ttggaatata	240
agttactcat	gagaaactcg	tgacttcttc	agcctaccca	aaccccgaca	tccgtgattg	300
catgaatctg	gccggagcct	acaacttgaa	ggccacagaag	ctctgacttt	accagctcat	360
gtccagcgac	aaacaagatc	taaccattgg	gcacctcgcc	ctacacatca	tggccctcac	420
ctctctctct	cgagaccctg	gggataaagt	atccattcta	caaaagacaa	tggagaactg	480
ggcacccttc	agcccccaag	ctgaagcatc	agccttctat	gggcccagtc	tagcgatctt	540
ggcactgtgc	cagaagaact	ctgaggcgac	cttggccgata	gccgttcgct	ttgccaagac	600
ctctgtggcc	aactctcttc	ccttcaatgt	agacacagga	gcaatggcaa	ccttggctct	660
gacctgtatg	tacaacaaga	tcctctgtagg	ttcagaggaa	gggtacagat	ccctgttttg	720
tcagggtacta	aaggataatt	tggagaaaat	cagcatgaag	atcaagata	atggcatcat	780
tggagacatc	tacagtactg	gcctcgccat	gcaggctctc	tctgtaacac	ctgagccatc	840
taaaaaggaa	tggaaactga	agaagactac	ggatatgata	ctcaatgaga	taaagcaggg	900
gaaatccacc	aaccccatgt	coattgtctc	aatcctccct	tcctgaaag	ccaagacata	960
cctgatgtgt	ccccagttca	cttgtagtcc	tgatcatgag	gtacaaacca	cttaccaccg	1020
caacccctggc	cctggcccaca	ctctctgcat	taacatcact	gtcatataca	ccataaataa	1080
ccagctgagg	gggtgtgagc	tgctcttcaa	cgagaccatc	aattgttagt	tgaaaagtgg	1140
gtcagtgtta	cttgttgttc	tagaggaagc	acagcgcaaa	aatcctatgt	tcaaatgtga	1200
aaccacaagt	acatcttggg	gccttctgtc	ctcttctatc	aacaatatcg	cggaaaatgt	1260
taatacacaag	acataactggt	agtttcttag	tgggttaaca	cctttgaatg	aagggtgtgc	1320
tgactacata	cccttcaacc	acgagcacat	ccagaccaat	tgcacacagt	actaacgaag	1380
aggtgggttc	agcttctatc	aaacatctcc	aaagatttgg	tgaaattttt	tccacttcat	1440
tttaaatcta	tgcataaaag	cgaatgcctg	tgatgtcacc	atatctctcg	taaaaacatg	1500
gagaaccact	atgtagaata	aaaatgcaaa	gttcaactgga	gtctcaacat	ctatgactca	1560
tgaataataa	attttcatct	tctc				1584

<210> 32
 <211> 1537
 <212> DNA
 <213> Homo sapiens

<400> 32						
gctctcatta	ccttctgccc	atcacttaat	aaatagccag	ccaattctac	aacattcttg	60
tacactgttg	gagagatgag	acagtcacac	cagctgcccc	tagtggggct	cttactgttt	120
ctcttttatc	caagccaact	atgcgagatt	tgtgaggtaa	gtgaagaaaa	ctacatccgc	180
ctaaaacctc	tgttgataac	aatgatccag	tcaaaactta	acaggggaa	cagcgctgtc	240
aatgttgtgt	tgtccctcaa	actgtttgga	atccagatcc	aaacctgat	gcaaaagatg	300
atccaacaaa	tcaaatataa	tgtgaaaagc	agattgtcag	atgtaagctc	gggagagctt	360

gcagggtctgc tcagtgctgc cctctgacaa aattaaagca ttgatggcct gtggacctgc 1860
 aaaaaa 1866

<210> 34
 <211> 2798
 <212> DNA
 <213> Homo sapiens

<400> 34
 gccctctccc acagcggagt ccaaacacagg cctaccagtc agttcttatt tctattgggt 60
 gtttccatgc tccaccatgt taagagctaa gaatcagctt tttttacttt cacctcatat 120
 cctgaggcag gtacaaagat catcaggctc caggctcata cagcaacgac tctacacca 180
 ccaacagccc cttcaccagc aatgggctgc cctggctaaa aagcagctga aaggcaaaaa 240
 cccagaagac ctaatatggc acaccccgga agggatctct ataaaacct tgtattccaa 300
 gagagatact atggacttac ctgaagaact tccaggagtg aagccattca cactgggacc 360
 atatcctacc atgtataact ttaggccctg gaccatccgc cagtatgtgt gttttagtag 420
 tbtggaagaa agcaataagt tctataagga caacatataag gctggctcagc agggattatc 480
 agttgctttt gatctggcga cacatcgttg ctagattcca gacaaccctc gacttcgttg 540
 tgatgttggg atggctggag ttgctattga cactgtggaa gataccaaaa tctcttttga 600
 tggaaattct tttagaaaaa tgtcagtttc catgactatg aatggagcag ttattccagt 660
 tcttgcaatt tttatagtta ctggagaaga acaaggtgta cctaaagaga aacttactgt 720
 taccatccaa aatgatatac taaagggaatt tatgtttcga aatacatata tttttctctc 780
 agaacctatc atgaaaatta ttgctgacat atttgaatat acagcaaaag acatgccaaa 840
 atttaattca atttcaatta gtggatacca tatgcaggaa gcaggggctg atgccattct 900
 ggagctggcc tatacttttag cagatggatt ggagtactct agaactggac tccaggctgt 960
 cctgacaatt gatgaatttg caccaagggt gtctttcttc tggggaattg gaatgaattt 1020
 ctatatggaa atagcaaga tgagagcttg tagaagactt tgggctcact taatagagaa 1080
 aatgtttcag cctaaaaact caaaatctct tcttctaaga gcacactgtc agacattctg 1140
 atgtgtcact actgagcagg atccctacaa taatatgttc cgtactgcga tagaagcaat 1200
 ggcagcagta tttggaggga ctcagttctt gcacacaaat tcttttgatg aagctttggg 1260
 tttgccaaact gtgaaaagtg ctgcaattgc caggaaacaca caaatcatca ttcaagaaga 1320
 atctgggatt cccaaagtgg ctgatccttg gggaggttct tacatgatgg aatgtctcac 1380
 aaatgatgtt tatgatctg ctttaaagct cattaatgaa attgaagaaa tgggtggaat 1440
 ggccaaagct gtactgtagg gaatacctaa acttgaatt gaagaatgtg ctgccgaag 1500
 acaagctaga atagattctg gttctgaagt aattgttggg gtaataaagt accagtttga 1560
 aaaaagaagac gctgtagaag ttctggcaat tgataatact tcaagtgcga acaggcagat 1620
 tgaaaaactt aagaagatca aatccagcag ggatcaagct ttggctgaac atgtcttgc 1680
 tgcataaacc gaattgtgct ctagcggaga tggaaaatct ctggctcttg cagtggatgc 1740
 atctcgggca agatgtacag tgggagaaat cacagatgcc ctgaaaaagg tattttgtga 1800
 acataaaggc aatgatcgaa tggtagtggt agcatatcgc caggaattttg gagaaagtaa 1860
 agagataaca tctgctatac agaggggttca taattctatg gaacgtgaag gtcgcagacc 1920
 tctgtctctt gtacgcaaaa tgggacaaga tggccatgac agagaggcaa aagttattgc 1980
 tacaggaatt gctgatctg gttttgatgt ggcacataggc cctcttttcc agtactctgc 2040
 tgaagtggcc cagcaggctg tggatcgga tgtgcatgtg gtgggctgaa gcacctctgc 2100
 tctgtgtcat aaaaacctag ttctgaaact catcaaaaga cttactccc ttggacggcc 2160
 agatattctt gtcattgtgt gagggtgat accacctcag gattatgaat tctgttttga 2220
 agttgtgtgt tccaatgtat ttggtcctgg gactcgaatt ccaaaggctg ccgttccagt 2280
 gcttgatgat attgagaagt gtttggaaaa gtagcagaaa tctgtataat atctctttt 2340
 tgttttagct ttgtctcaaa atattatttt agttatgac aaagaagaga gtaaaagcat 2400
 gcttccaatt taatttcaat acctgatatt tactttcttt gaaagcttta ctttaaaata 2460
 ccttactatt aggcctgggt tcatgtcata agtatgtaca tacagtttca cctcaaaaaa 2520
 aaaaaaaaat cctcaaaaac tctctatact ctctataaca atactttatc aagaactctg 2580
 gacaattgta ttatttttaa aaatcatgtg gatgtattta ttagaattgt tctataaat 2640
 ccttttcaat tttatattaa gaattaaact gtacctaaaa aaactctgac tattccatt 2700
 tctcagttta gcattacatt gtcttgagca ccagaaaaata aaatccatat attaatataa 2760
 acctatcttg aaaaaaaaaa aaaaaaaaaa aaaaaaaa 2798

<210> 35
 <211> 1637
 <212> DNA
 <213> Homo sapiens

<400> 35
 aagaactggc ctgtacattt tcaaggaatt cttgagaggt tcttgagag attctgggag 60
 ccaaacactc cattgggagc ctgactgttt tagaagaacaa cttgtaatgg agccttcac 120
 tcttgagctg ccggctgacg cagtgccagc cattgcccgt gaactcaaat gccacccaac 180
 ggatgagagg gtggctctcc acctagatga ggaagataag ctgagggcact tcaggggagt 240
 cttttatatt cccaaaatac aggatctgcc tccagttgat ttatcattag tgaataaaga 300
 tgaataatgcc atctatttct tgggaaatct ctttggcctt caaccataaa acattttaa 360
 atatcttgaa gaagaactag ataagtggcg caaaatagca gccatgggtc atgaagtggg 420
 gaagcgctct tggattacag gagatgagc tattgtaggc cttatgaagg acattgtagg 480
 agccaatgag aagaataatag ccctaataatg tgctttgact gtaaatttac atcttctaat 540
 gtatcatctt ttttaagccta cgccaaaacg atataaaatt cttctagaag ccaaagccct 600
 ccctctctgat cattatgcta ttgagtccaa actacaactt caccgactta acattgaaga 660
 aagtatgcgg atgataaagc caagagaggg ggaagaaacc ttaagaatag aggatattct 720
 tgaagtaatt gagaaggagag gagactcaat tgcagtgatc ctgttcagtg ggggtgcatt 780
 ttacactgga cagcacttta atattcctgc catcacaaaa gctggacaag cgaaggggtg 840
 ttatgttgcc tttgatctag cacatgcagt tggaaatgtt gaactctact tacatgactg 900
 gggagttgat ttgtcctgct ggtgttccca caagtattta aatgcaggag caggaggaat 960
 tgctgggtgcc ttcattcatg aaaagcatgc ccatacgatt aaacctgcatt tagtgggatg 1020
 ggttggccat gaactcagca ccagattttaa gatggataac aaactgcagt taactccctg 1080
 gtctctggga ttccgaattt caaatccctc cattttgttg gtctgttctt tgcattgtag 1140
 ttttagagatc ttttaagcag cgacaatgaa ggcattgcgg aaaaaatctg ttttgctaac 1200
 tggctctctg gaattcattg tcaagcataa ctatggcaaa gataaagcag caaccaagaa 1260
 accagtctgt aacataatta ctccgtctca ttagtagagg cgggggtgccc agctacaact 1320
 aacattttct gtccccaaac aagatgtttt ccaagaacta gaaaagaagag gagtgggttt 1380
 tgacaagcgg aatccaaatt gcattcgagt ggctccagtt cctctctata attctttcca 1440
 tgatgtttat aaattttacca atctgctcac ttctataact gactctcgag aaacaaaaaa 1500
 tttagcagtt ttcttagaac aacttaagca aattatactg aaagctcagt tgggttatttc 1560
 agtatatttc gatttttaac tattgaaagt atgtcaccat tgaccacatg taactaacia 1620
 taataaatat accttac 1637

<210> 36
 <211> 1908
 <212> DNA
 <213> Homo sapiens

<400> 36
 gaattcatga aaacgtagct cgtcctcaaa aaaaacagaa gaggagtaac cattttaagg 60
 gagaataata tacgaaggag acaagatttt gaagcaccga agctgccacc tacattaaaa 120
 caccgttagt ggctaaacac cagtcttcaa tgcctctcca cagcctcact ctgaaaaata 180
 ctgtgcaggt gaccacaagt aggggtcacc cttgggcttt tctgttgcca gtatctctgg 240
 tttagaaaaa acaaacctga cttattgcgt tgaagacgag caacaggaaag gactccatga 300
 tttagtcact ctataccatc ctaagaaact ttatccacc aaactgtatt tcagacttta 360
 taatctaaac tacaaaaagt gttcactggg gaactgcaca atatgactgc ttttaaccgt 420
 agtgatttca aatattgagc catgctgttg cagtctttaa aactggagag ctaagggcag 480
 cttctctcta gtcaccaaat ccagcacttt tttaaaaaat cagtaaaact ctctgaccac 540
 caaggaaaaa aaaaaaggat ggaggtttaa agacgcaccc cttgcccaaa agccccctca 600
 tcagaaattgg agtcaggaga cctgagttcc tgcctcaggc ctgcatttaa aaactgcact 660
 aacctttggc tatctctcca aacggaagta ctaaaacctc agcgccttca ccaatttcta 720
 gccccggtgt ggctcttccc accttccctt tcttcagccc gcccttctct cctcagcccc 780
 tatcatcggt cggagggtccc ccgcctccgc ccgccttacc cacaagcccc gccccccagc 840
 ccccgatggc cctgcccagt cccagacaga acctactacg tgcggcgcca gctcggtgca 900
 gaaggcgggc gctggggggc ctgcggccgc tgcagcgcag ggtccacctg gtcgggtgca 960

```

cctgtggagg aggaggtgga ttccaggctt cccgtagact ggaagaatcg gctcaaaacc 1020
gctgtccctcg caggggctga gctggaggca gcgaggccgc ccgacgcagg ctcccggcga 1080
gacatggcag ggcaaggatg gcagcccgcg ggcagggccc ggcgaggagc gcgaacccgc 1140
ggccgcagtt cccaggcgctc tgcgggcgcg agcacgccgc gaccctgcgt gcgccggggc 1200
ggggggggcg ggccctgcct gcacaaatag ggacgagggg gcggggcgcg cacaatttcg 1260
cgccaaactt gaccgcgcgt tctgctgtaa cgacggggct cggaggctct cccgctgctg 1320
tcatgggttg ttgcctaaac tgcctcgtcg ctgtgtccca gaacatgggc atcggaaga 1380
acggggacct gccctggcca ccgctcaggt atctgccggg ccggggcgat gggacccaaa 1440
cggggcgcag ctgcccacgg tcgggggtacc tgggggggac gcgccggcg actccggcg 1500
agaggatggg gccagacttg cggctctgcg tggcaggaag ggtggggcgc actggattcc 1560
ctttttctgc tgcgggggag gccagttgc tgatttctgc ccgattctg ctgcccggtg 1620
aggctcttgc ctgcgggcgc ctgcgccagg gcaaaagtcc agccctggag aaacacctc 1680
acccctaccc acacgcctcc gtttgtcagg tgccttagag ctgcagccca agggataatg 1740
tttcagtaga cgctgtttct ctaacttgta ggaatgaatt cagatatttc cagagaatga 1800
ccacaacctc ttcagtagaa ggtaatgtgg gattaagtag ggtctgtctt gatgaagttt 1860
accagtgcac atgttagtta aatggaaagt ttcccggtgt aatctggg 1908

```

```

<210> 37
<211> 30
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence:primer

```

```

<400> 37
cccacggtcg gggtggccga ctcccggcga 30

```

```

<210> 38
<211> 21
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence:primer

```

```

<400> 38
ctaaactgca tcgtcgtgt g 21

```

```

<210> 39
<211> 19
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence:primer

```

```

<400> 39
aaaaggggaa tccagtcgg 19

```

```

<210> 40
<211> 19
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence:PCR product

```

0577266-0577266

19

[illegible]

#400> 42						
ctgcagcgcc	agggtcacc	tggtcgctg	cacctgtgga	ggaggagggtg	gatttcagrg	60
tctccgtata	ctggaagat	cggctcacc	ccgcttgct	cgcaggggt	gagctggagg	120
cacgcagaga	gccgcagca	ggcttccgg	gagacatg	agggcaagga	tgcgcagccg	180
ggcgaggggc	ccgcgagga	gcgcgaacc	gcggccgac	ttccagagc	tctggagggc	240
ggagcagcgg	gcgacccctg	gtgcgcggg	cggggggggg	ggggctctgc	ctcgcaaaat	300
agggacaggg	ggggcggggc	gccacaatt	gcgcgcaaac	tgcacgcgc	gtctctgtgt	360
aacgacggcc	ctccggagtc	ctccctctg	tgtcatgtgt	gggtctgtaa	actgcatcgt	420
ctgctgttgc	cagaaacatg	catctcgcaa	gaacggggac	ctgcctctgc	ccacgcctag	480
gtatctccgc	ggcgggggcg	atggggacca	aacgggcgca	ggctgccca	gggtgggggt	540
ggcgactcgc	ggcgagagga	tggggcaga	cttgcctgt	gcgctggcag	gaaggtgggt	600
cccgactagg	tcccccttt	ctgtctgcg	ggagggccag	tgtctgatt	ctgcgccgat	660
tctgtctgcc	ggtaggtgt	tgtccctgc	ggccctctgc	ccaggcgaaa	gtcccaagc	720
tggtagaaac	acctcaccgc	taccaccagc	gctccgtttg	ctaggtctgc	tagatcgata	780
gcccaagaaa	taattgttgc	agtaacgtg	tttctctaac	tcttgagga	taagtacgat	840
atttccagag	aaatgacaca	acctcttcag	tagaagaatg	tgtgggatta	agtaggaatc	900
tgtctgatga	agtttaccag	tccaatgtg	aggttaatgt	aaagtgttcc	gtgttaatct	960
ggcgaccttt	ctcttattat	ggatctgat	gatctgtatg	cagttcccaa	ggttcatta	1020
ccattattta	aaaatttttg	tcttagaata	tttatgtatg	tcgaaccacg	agcaaaatt	1080

caggcatggg gcagaattgg caactgggtg gaggtctcgg tggaggttag cactccgaaa 1140
 ggaaacaga gtatgacctt ggaacagctg ctggaagaga taaggcctga acaagggcag 1200
 tggagaagag agggtaaaaa ttttttaagg ttacatgacc ctggattttt gagatc 1256

<210> 43

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:PCR product

<400> 43

gctgcccacg gtccgggtac ctggggcgga cgcgccaggc cgactcccg cgaga 55

<210> 44

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:PCR product

<400> 44

gctgcccacg gtccgggtgg ccgactcccg gcgaga 36

<210> 45

<211> 1273

<212> DNA

<213> Homo sapiens

<400> 45

ctgcagcgca gggccacct ggctcgctgc acctgtggag gaggaggtgg atttcaggct 60
 tcccgtagac tggagaatc ggctcaaac cgcttgccctc gcaggggctg agctggaggc 120
 agcgaggccg cccgacgcag gcttccggcg agacatggca gggcaaggat ggcagcccg 180
 cggcaggggc cggcgaggag cgcgaacccg cggccgcagt tcccaggcgt ctgcggggcg 240
 gagcacgccc cgaccctgcg tgcgccgggg cggggggggc gggcctcgcc tgcacaaata 300
 gggacgaggg ggcggggcg ccacaatttc gcgccaaact tgaccgcgcg ttctgtgtga 360
 acgagcgggc tcggaggtcc tcccgcgtgt gtcatgtttg gtctgctaaa ctgcatcgtc 420
 gctgtgtccc agaacaatgg catcggaag aacggggacc tgccttgccc accgctcagg 480
 tatctgcgcy gccggggcga tgggacccaa acgggcgcag gctgccacg gtcgggggtac 540
 ctggggcgga cgcgcgggcc gactccggcg gagaggatgg ggcacagact gcggtctgcg 600
 ctggcaggaa ggggtggccc gactggattc cctttttctg ctgcgcggga gcccagttg 660
 ctgattttct cccgatttct gctgcccggt gagggttttg cctcgccgcy ccctcgccca 720
 gggcaaaatc ccagccctgg agaaaaacac tacccttac ccacagcgct ccgtttgtca 780
 ggtgccttag agctcgagcc caagggataa tgtttcgagt aacgctgttt ctctaaacttg 840
 taggaatgaa ttcagatatt tccagagaat gaccacaacc tcttcagtag aaggtaatgt 900
 gggattaaat agggctctat ttgatgaagt ttaccagtcg aaatgttagt taaatggaaa 960
 gtcttccgtg ttaattctgg accctttctc ttattatgga tctgtatgat ctgtatgcag 1020
 ttcccaaggt tcatttacca ttattataaaa atttttgtct tagaaaatttt atgtatgtca 1080
 cgcacgagc aaattatcag gcatggggca gaattggcaa ctgggtggag gcttcggtgg 1140
 aggttagcac tccgaagga aaacagagta ggcctttgga acagctgctg gaagagataa 1200
 ggcctgaaca agggcagttg agaagagagg gtaaaaaatt tttaaggtta catgaccctg 1260
 gattttggag atc 1273

<210> 46

<211> 18

